

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa de Pós-graduação em Bioinformática

Priscilla Alves de Aquino

**“DIVERSIDADE MOLECULAR DE FOSFOLIPASES D DA PEÇONHA
DA ARANHA *Loxosceles Laeta* PERUANA REVELADA POR
SEQUENCIAMENTO DE NOVA GERAÇÃO E ANÁLISE
TRANSCRIPTÔMICA”**

Belo Horizonte

2016

Priscilla Alves de Aquino

**“DIVERSIDADE MOLECULAR DE FOSFOLIPASES D DA PEÇONHA
DA ARANHA *Loxosceles Laeta* PERUANA REVELADA POR
SEQUENCIAMENTO DE NOVA GERAÇÃO E ANÁLISE
TRANSCRIPTÔMICA”**

Dissertação apresentada ao curso de Pós-Graduação em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito para obtenção de título de Mestre.

**Orientador: Prof. Dr. Carlos Delfin Chávez
Olórtgui**

**Coorientadora: Dra. Ana Paula Vimieiro
Martins**

Belo Horizonte

2016

043 Aquino, Priscilla Alves de.
Diversidade molecular de Fosfolipases D da peçonha da aranha *Loxosceles laeta* peruana revelada por sequenciamento de nova geração e análise transcriptômica [manuscrito] / Priscilla Alves de Aquino. – 2016.
91 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Carlos Delfín Chávez Olórtegui. Coorientadora: Dra. Ana Paula Vimieiro Martins.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Transcriptoma. 3. Fosfolipase D. 4. RNA-Seq. 5. Aranha Marrom Reclusa. I. Olórtegui, Carlos Delfín Chávez. II. Martins, Ana Paula Vimieiro. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DA DEFESA DE DISSERTAÇÃO

Priscilla Alves de Aquino

25/2016
entrada
2º/2014
CPF:
067.825.176-20

Às quatorze horas do dia **31 de agosto de 2016**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Diversidade Molecular de Fosfolipases-D da Peçonha da Aranha *Loxosceles laeta* Peruana revelada por sequenciamento de nova geração e Análise Transcriptômica**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, o Presidente da Comissão, **Dr. Carlos Delfin Chávez Olórtegui**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Dr. Carlos Delfin Chávez Olórtegui	UFMG	455731316-72	APROVADA
Dr. Evanguedes Kalapothakis	UFMG	49430742604	APROVADA
Dra. Liza Figueiredo Felicori Vilela	UFMG	05000749697	APROVADA
Dr. José Miguel Ortega	UFMG	059501265-07	Aprovada

Pelas indicações, a candidata foi considerada: _____
O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.
Belo Horizonte, 31 de agosto de 2016.

Dr. Carlos Delfin Chávez Olórtegui - Orientador Carlos Chávez

Dr. Evanguedes Kalapothakis Evanguedes Kalapothakis

Dra. Liza Figueiredo Felicori Vilela Liza Felicori

Dr. José Miguel Ortega José Miguel Ortega

Este trabalho foi realizado com o apoio das seguintes instituições:

- Programa de Pós-Graduação em Bioinformática - ICB - UFMG

- Departamento de Bioquímica e Imunologia – ICB – UFMG;

- Fundação de Amparo à Pesquisa de Minas Gerais – FAPEMIG;

- Conselho Nacional de Pesquisa e Desenvolvimento– CNPq;

- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES

Dedico esse trabalho à minha família.

AGRADECIMENTOS

Ao programa de Pós-Graduação em Bioinformática e ao departamento de Bioquímica pela oportunidade e apoio para o desenvolvimento desse trabalho e pela oportunidade de aprendizado e convívio com grandes pesquisadores do Brasil.

Às agências de fomento CAPES, CNPq e FAPEMIG pelo financiamento desse projeto e pela bolsa de mestrado.

À banca por ter aceitado o convite de participar da arguição da minha dissertação e pela disponibilidade, tenho certeza que a presença de vocês trará grandes contribuições a esse trabalho.

Ao professor Carlos Delfín Chávez Olórtegui pelo apoio durante todo o tempo de vivência, com grande competência, sensibilidade e experiência. Aos membros do laboratório de Imunoquímica de Proteínas e colaboradores, especialmente a Clara Guerra, pelas colaborações, incentivo e apoio.

Ao professor Evanguedes Kalapotakis e aos membros do laboratório de Biotecnologia e Marcadores Moleculares, em especial ao Arthur Dantas, pelo auxílio e por compartilhar sua experiência na análise transcriptômica. Agradeço também especialmente à Ana Paula Vimieiro Martins, pela orientação, paciência e indispensável colaboração durante esse trabalho, que sempre que necessário me indicou a direção certa a seguir para superar as adversidades surgidas no percurso. Ana, você foi essencial nesse período tão importante contribuindo muito para que essa etapa da minha vida fosse passível de ser concretizada. Muito Obrigada!

Ao professor José Miguel Ortega e equipe do laboratório de Biodados, pela disposição a auxiliar e esclarecer dúvidas, sempre com profissionalismo, competência, bom humor e incentivo.

Ao professor e coordenador Vasco Ariston, pelo voto de confiança e oportunidade, bem como pelas sugestões e direcionamento sempre oportunos e importantes para o futuro acadêmico.

À todos os amigos e colegas da pós-graduação, em especial à Michele Araújo pela disponibilidade em compartilhar seus conhecimentos e ao Carlos Alberto Xavier pela paciência, prestabilidade e companheirismo ao longo do curso, o qual se tornou uma pessoa de confiança com quem eu sempre pude e posso contar. Que a força esteja com você!

Aos meus queridos amigos de convívio, pelo apoio, pela paciência e compreensão das minhas ausências e stress durante esse período. Um agradecimento especial aos grandes amigos Luisa Rodrigues e Roberto Sakamoto, pelo carinho e apoio incondicional.

À toda minha família pela confiança e toda ajuda durante esse período, em especial aos meus pais e irmãos que acompanharam mais de perto esse momento.

RESUMO

Este estudo objetiva a análise por ferramentas bioinformáticas de transcritos da glândula de peçonha da aranha-marrom *Loxosceles laeta*, para a ampliação do banco de dados de fosfolipases D. As sequências encontradas através do sequenciamento de nova geração RNA-seq e montagem do transcriptoma pelo programa Trinity, foram submetidas à anotação gênica para que fossem identificadas por meio de buscas de similaridade contra diferentes bancos de dados. Em seguida, foram classificadas em categorias funcionais e avaliadas quanto a sua função biológica a partir de análise filogenética. A identificação correlacionada aos compostos da peçonha favorece melhor compreensão dos processos fisiopatológicos do envenenamento, bem como ferramentas para melhorias nos métodos de profilaxia e tratamento para o Loxoscelismo. Nesta dissertação objetivamos nossos estudos transcriptômicos nas proteínas da peçonha de *L. laeta* responsáveis pela principal atividade tóxica do composto, que são as Fosfolipases D, esfingomielinases D ou proteínas dermonecróticas. Nesse trabalho, foram descritas 19 sequências de fosfolipases D potencialmente encontradas na peçonha de *L. laeta*. A partir da análise e descrição da expressão das fosfolipases D da peçonha e interatividade molecular, torna-se possível o aperfeiçoamento de ferramentas para o estudo dos mecanismos de ação da atividade dermonecrótica dessas enzimas. Além disso, o estudo aprofundado da expressão dessas enzimas na peçonha e sua constituição favorecem futuros estudos sobre a filogenia de suas proteínas.

Palavras-chave: aranha-marrom ; *Loxosceles laeta* ; Transcriptoma ; Fosfolipase D ; RNA-Seq

ABSTRACT

This study aims to analyze, by bioinformatic tools, the transcripts of the brown spider venom gland *Loxosceles laeta* to expand the phospholipase D database. Sequences found through the new generation of sequencing RNA-seq and transcriptome assembly of the program Trinity, were subjected to gene annotation that so they could be identified through searches of similarity against different databases. Then they were classified into functional categories and had their biological function evaluated from a phylogenetic standpoint. The identification correlated to the venom of compounds promotes better understanding of the pathophysiological processes of envenomation cases, as well as tools for improvements in methods of prophylaxis and treatment for loxoscelism. This dissertation aimed to describe the venom proteins responsible for its major toxic activities, which are the Phospholipase D, Sphingomyelinase D or Dermonecrotic Proteins. In this study, 19 putative Phospholipase D sequences were described from the venom gland of the spider *L. laeta*. From the analysis and description of the Phospholipases D expression from the venom and their molecular interaction, it is possible to improve tools for the study of mechanisms of action of dermonecrotic activity of these enzymes. In addition, the in-depth study of the expression of these enzymes favor future studies on the phylogeny of their proteins.

Keywords: Brown Spider ; *Loxosceles laeta* ; Transcriptome ; Phospholipase D ; RNA-seq

LISTA DE FIGURAS

Figura 1 Aranha fêmea do gênero <i>Loxosceles</i> também conhecida como “aranha-marrom”.....	16
Figura 2 Mapa geográfico com as espécies mais encontradas do gênero <i>Loxosceles</i> em cada continente.	17
Figura 3 Espécimes adultos de <i>Loxosceles laeta</i> coletados no município do Rio de Janeiro, Brasil.	18
Figura 4 Visão do dorso e posicionamento, demarcado em vermelho, da glândula da peçonha da aranha-marrom no cefalotórax.	18
Figura 5 Evolução de lesão ocasionada por picada de <i>Loxosceles</i>	20
Figura 6 Distribuição da porcentagem do nível de expressão de cada grupo de toxinas e enzimas produzidas na glândula de peçonha da aranha <i>L. intermedia</i>	25
Figura 7 Esquema mostrando metodologia de RNA-Seq.	29
Figura 8 Visão geral do Trinity.	36
Figura 9 Valores de qualidade Q por base ao longo das reads antes da trimagem..	39
Figura 10 Valores de qualidade Q por base ao longo das reads após trimagem.....	41
Figura 11 Abundância de fosfolipases no transcriptoma da glândula de peçonha de <i>Loxosceles laeta</i>	45
Figura 12 Alinhamento entre as sequências protéicas dos contigs do transcriptoma de <i>Loxosceles laeta</i>	55
Figura 13 Alinhamento entre as sequências protéicas dos contigs do transcriptoma de <i>Loxosceles laeta</i> peruana e sequencias do NCBI de <i>Loxosceles laeta</i> brasileira.	57
Figura 14 Cladograma do alinhamento múltiplo efetuado de similaridade com as enzimas de fosfolipases – D encontradas na peçonha de <i>L.laeta</i> contra outras proteínas do NCBI com melhores hits entre o alinhamento do Blastx.....	59

LISTA DE QUADROS

Quadro 1 Classificação quanto à gravidade, quadro clínico e tratamento dos acidentes por aranhas-marrom	21
Quadro 2 Opções do programa Prinseq-lite utilizados para trimagem das sequências de baixa qualidade e suas respectivas funções de acordo com o manual do usuário....	34
Quadro 3 Opções do programa Trinity, e suas funções correspondentes, utilizadas para a montagem de novo das reads.	37

LISTA DE TABELAS

Tabela 1. Vantagens do método RNA-Seq comparado aos outros métodos de sequenciamento.	28
Tabela 2 Valores de Phred e correspondente probabilidade de erro de nomeação das bases.....	33
Tabela 3 Relação quantitativa de sequências geradas antes e após trimagem.....	42
Tabela 4 Contigs; Peptídeos do NCBI com melhor alinhamento; peptídeo sinal em negrito e sublinhado; sequência madura após o peptídeo sinal; e-value; P.I. Ponto Isoelétrico; M.M. Massa Relativa	50

LISTA DE ABREVIATURAS

°C	Graus Celsius
BLAST	Basic Local Alignment Search Tool
cDNA	DNA complementar
DNA	Ácido desoxirribonucleico
ELISA	Enzyme linked immunosorbent assay
Gb	Gigabases
kb	Quilobase
kDa	Quilodalton
LoxTox	“Loxosceles’ toxin”
M	Molar
M.M.	Massa Molecular
mM	Milimolar
mg	Miligrama
mL	Mililitro
µg	Micrograma
µL	Microlitro
mRNA	RNA mensageiro
NCBI	Centro Nacional de Biotecnologia
NGS	Next Generation Sequencing
OMS	Organização Mundial de Saúde
ORF	Open reading frame (fase de leitura)
pb	Pares de bases
PCR	Reação em cadeia da polimerase
PLD	Fosfolipase -D
pH	Potencial Hidrogenionte
RNA	Ácido ribonucléico
RNA-Seq	RNA sequencing
SicTox	“Sicarius’ Toxin”
Swiss prot	Protein knowledgebase of Swiss Institute

SUMÁRIO

1	INTRODUÇÃO	15
1.1	A Aranha.....	15
1.2	O Loxoscelismo	19
1.3	Epidemiologia do Loxoscelismo	22
1.4	A Peçonha da <i>Loxosceles</i>	22
1.5	A Bioinformática na Transcriptômica	26
1.5.1	Introdução ao estudo computacional de transcriptomas.....	26
1.5.2	RNA-Seq e a montagem <i>de novo</i>	28
2	JUSTIFICATIVA	30
3	OBJETIVOS	30
3.1	Objetivo Geral	30
3.2	Objetivos Específicos	31
4	METODOLOGIA	31
4.1	Extração de peçonha de <i>L. laeta</i>	31
4.2	Construção da biblioteca para sequenciamento de nova geração	32
4.3	Sequenciamento de Nova Geração (NGS) – MiSeq Illumina.....	32
4.4	Pré-processamento das <i>reads</i>	33
4.5	Montagem dos <i>contigs</i>	35
4.6	Anotação das sequências relacionadas a fosfolipases	37
5	RESULTADOS E DISCUSSÃO	39
5.1	Construção da biblioteca de cDNA	39
5.2	Pré-processamento das sequências.....	40
5.3	Montagem <i>de novo</i>	42
5.4	Visão geral das fosfolipases no transcriptoma da peçonha Loxoscélica	43
5.5	Anotação de Isoformas de Fosfolipases - D da peçonha de <i>L. laeta</i>	45
5.5.1	Identificação de sequências com similaridade às proteínas do banco de dados NCBI....	45
5.5.2	Processamento dos <i>contigs</i> selecionados: tradução, identificação do peptídeo sinal, do ponto isoelétrico (P.I.) e massa molecular (m.m)	46
5.5.3	Alinhamento entre as sequências de proteínas completas do transcriptoma da Aranha-marrom peruana e proteínas previamente descritas	52
5.5.4	Visão geral da Filogenia da Aranha-marrom peruana.....	58
6	CONCLUSÕES	60
7	PERSPECTIVAS	60
8	REFERÊNCIAS BIBLIOGRÁFICAS	62
9	ANEXOS	74
9.1	Anexo 1 – Fasta com as sequências do cDNA de <i>L. Laeta</i> correspondentes aos contigs obtidos através da montagem de novo de RNA-Seq.	74
9.2	Anexo 2 – Sequências traduzidas dos contigs obtidos através de sequenciamento RNA-seq, através da ferramenta on line Expsy Translate.	82
9.3	Anexo 3 - Alinhamento entre proteínas do transcriptoma de <i>L.laeta</i> peruana e PLDs, de diversas espécies de aranhas do gênero <i>Loxosceles</i> e <i>Sicarius</i> , do banco de dados do NCBI.	84

1 INTRODUÇÃO

1.1 A Aranha

Os artrópodes (do grego arthros: articulado e podos: p.s, patas, apêndices) constituem o filo mais expressivo, estimados em 70% da biodiversidade de todas as espécies descritas atualmente (GRIMALDI e ENGEL, 2005). Em geral, são encontrados em todos os tipos de ambientes, possivelmente pela sua anatomia, sendo que uma de suas características mais representativas é a presença de exoesqueleto composto de quitina (BRUSCA E BRUSCA, 2007). Dentro do filo Arthropoda, os aracnídeos são tratados como uma classe (Arachnida) muito significativa e representam o segundo maior grupo do Reino Animal depois do grupo dos insetos, sendo considerada a mais importante dentro do subfilo Chelicerata (FOELIX, 2011). Os organismos mais representativos da classe Arachnida são as aranhas (Ordem Araneae), os escorpiões (Ordem Scorpiones), os opiliaes (Ordem Opiliones), os ácaros e os carrapatos (Ordem Acari).

A ordem Araneae, é constituída por aproximadamente 46.000 espécies, divididas em cerca de 110 famílias e 3.750 gêneros (PLATNICK & RAVEN, 2016). No Brasil, são encontrados três gêneros de aranhas que apresentam maior importância médica: *Phoneutria*, *Loxosceles* e *Latrodectus*. As aranhas são em sua maioria noturnas, ocupam quase todos os ambientes naturais e, devido à sua alta capacidade de dispersão, podem ser encontradas em grande número nas áreas urbanas, representando uma característica adaptativa importante (SANTOS et al., 2007). As aranhas do gênero *Loxosceles* (NICOLET, 1849) (Figura 1), são próprias de centros urbanos e de origem no continente Africano. Possuem 114 espécies diferentes, e estão dispersas pelos cinco continentes com relatos de acidentes com humanos nas Américas, região do Mediterrâneo, Austrália, Índias Ocidentais e África (Figura 2) (BRIGNOLI, 1976; GERTSCH, 1967; GERTSCH E ENNIK, 1983; GERSCHMAN AND SCHIAPPELLI, 1979; NEWLANDS E ATKINSON, 1988; WANG, 1994).

Na América do Sul, a *L. laeta* é encontrada em diversas regiões do sul do Brasil, por toda extensão do território argentino, com elevada incidência no Chile e finalmente com alto número de relatos em toda região do Peru (DE ROODT et AL., 2002; MALAQUE ET AL., 2002; MANRIQUEZ E SILVA, 2009; MAGUIÑA-VARGAS ET AL., 2004). No território peruano, principalmente na costa do país, as espécies de maior importância para saúde pública são a *L. laeta* e *L. Rufipes*, devido ao maior número de relatos de acidentes loxoscélicos (SANABRIA E ZAVALETA, 1997).



Figura 1 Aranha fêmea do gênero *Loxosceles* também conhecida como “aranha-marrom”.

Fonte: <http://www.arachnoserver.org/toxincard.html?id=AS000273>

No Brasil existem oito espécies de *Loxosceles*, sendo quatro endêmicas no país; a saber: *L. similis*, *L. gaucho*, *L. amazonica* e *L. puortoi*; e, quatro de países adjacentes: *L. laeta*, *L. intermedia*, *L. hirsuta* e *L. adelaida*.

Das espécies supracitadas, três interessam especialmente à medicina, porquanto causadoras de maior número de casos de acidentes no país: *L. intermedia*, *L. gaucho* e *L. laeta*. Estas são também as espécies mais encontradas no país, principalmente na região sul (Paraná e Santa Catarina) e região sudeste (São Paulo e Minas Gerais) (SAMS et al., 2001; FISHER, 2005; MINISTÉRIO DA SAÚDE, 2012; DA SILVA, 2004; PLATNICK, 2016).

As *Loxosceles* têm hábito noturno, preferência por lugares secos, escuros e quentes, onde possam construir teias com formato irregular remetendo a fios de algodão desfiados. Abrigam-se em depressões de barrancos, sob pedaços de madeira e pedras, entulhos, cascas de árvores e cavernas. No ambiente residencial, podem ser encontradas atrás de móveis, assoalho das casas, tijolos, fendas de paredes, telhas e entulhos em geral (FUTRELL, 1992; GERSTCH, 1967; LUCAS, 1988). São consideradas pouco agressivas para o homem, somente atacando quando se sentem ameaçadas. Quanto a seus hábitos alimentares, capturam pequenos invertebrados através da inoculação da peçonha para posterior digestão. Algumas espécies, como a *L. reclusa*, preferem as presas mortas às vivas e no geral resistem por meses sem água ou comida (FUTRELL, 1992; CARDOSO et al, 1988; FISCHER, 1996; DIAZ, 2004; RIBEIRO ET AL, 1993).



Figura 2 Mapa geográfico com as espécies mais encontradas do gênero *Loxosceles* em cada continente. Baseado em BRIGNOLI, 1976; GERSCHMAN E SCHIAPPELLI, 1979; GERTSCH, 1967; GERTSCH E ENNIK, 1983; NEWLANDS E ATKINSON, 1988; WANG, 1994.

Também conhecidas popularmente como “aranhas marrons” (VETTER AND VISSCHER, 1998) ou aranhas-violinos, apresentam o corpo de cor que varia entre marrom-amarelada (*L. laeta*) a marrom escuro (*L. gaucho*), com singelas variações entre as espécies, e possuem uma mancha escura na forma do instrumento. As aranhas deste gênero têm o porte pequeno, comprimento variável de 1,0 a 4,0 cm e apresentam três pares de olhos. As pernas apresentam lateralidade (Fig.1), e inclinação para frente em posição normal (CAMERON, 2005; FUTRELL, 1992; GERSTCH, 1967; LUCAS, 1988; DA SILVA et al., 2004). Os indivíduos dessas espécies possuem características diferentes entre machos e fêmeas, como por exemplo o tamanho do abdômen (Fig. 3). Dentre as espécies, os indivíduos de ambos os sexos possuem caráter peçonhento, contudo, a fêmea produz uma quantidade maior de peçonha, cuja atividade dermonecrótica é mais eficaz quando comparada à do macho (OLIVEIRA et al., 1999).

Nesses animais, a glândula da peçonha está localizada na cavidade do cefalotórax e apresenta estrutura bulbar, pareada, contendo dois ductos secretores que saem no aparelho inoculador, identificado como um par de quelíceras, como pode ser visto na figura 4. Essa glândula é composta por músculo estriado que se dispõe em duas camadas, uma na parte interna e a outra na externa (GAJARDO-TOBAR, 1966; FOIL et al, 1979; SANTOS et Al., 2000).

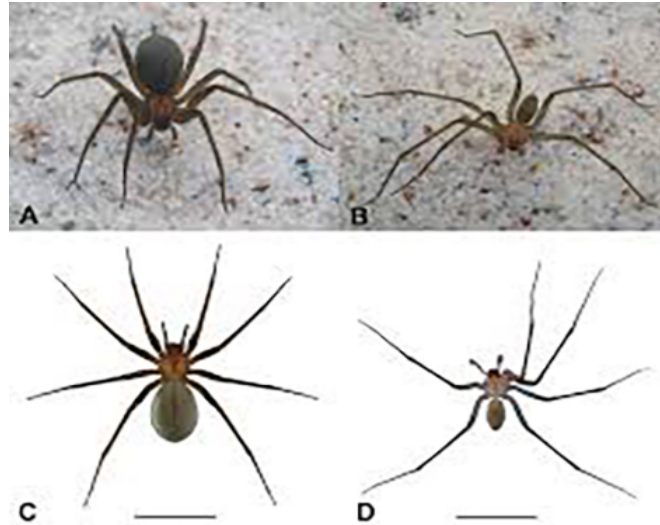


Figura 3 Espécimes adultos de *Loxosceles laeta* coletados no município do Rio de Janeiro, Brasil. A) Fêmea em vida B) Macho em vida C) Fêmea preservada D) Macho preservado. Fonte: SILVEIRA, ADRIANO LIMA. 2009

As vesículas secretoras que contêm a peçonha são resultado de projeções ao lúmen devido à disposição das células epiteliais lado a lado. Uma matriz colagenosa rica em fluido com glicoproteínas, resíduos sulfatados glicosaminoglicanos, laminina e entactina são encontrados na lâmina basal que está entre as células epiteliais secretoras e as células musculares das glândulas (GAJARDO-TOBAR, 1966; FOIL; COONS; NORMENT, 1979; SANTOS et Al., 2000).



Figura 4 Visão do dorso e posicionamento, demarcado em vermelho, da glândula da peçonha da aranha-marrom no cefalotórax.

Fonte:

http://www.pacoalarcon.com/aracnidos/details.php?image_id=172&mode=search

1.2 O Loxoscelismo

O conjunto de sintomas e sinais clínicos decorrentes do envenenamento causado pela picada das aranhas do gênero *Loxosceles* é denominado Loxoscelismo. O Loxoscelismo consiste na forma mais grave de araneísmo no Brasil e representa 50% dos casos em que a aranha causadora é identificada (MINISTÉRIO DA SAÚDE, 2012; SCHENONE et al., 1989; SEZERINO et al., 1998; DA SILVA, 2002). Em geral, o acidente loxoscélico pode ser definido como: um quadro clínico (FUTRELL, 1992) acarretando o desenvolvimento de duas formas clínicas, classificadas de acordo com um quadro cutâneo ou dermonecrótico, (84 - 97% dos casos); ou um quadro sistêmico (3 - 16% dos casos) (BARBARO et al., 1992; Ministério da Saúde, 2012). A gravidade dos casos está diretamente relacionada com a concentração da peçonha, a espécie do gênero, o sexo do animal, o estágio de desenvolvimento e a quantidade de peçonha inoculada (OLIVEIRA et al., 1999; MINISTÉRIO DA SAÚDE, 2012; SEZERINO et al., 1998). A princípio, a picada das *Loxosceles* tem caráter indolor e os sintomas aparecem (Gremski et al., 2014).

O Loxoscelismo cutâneo é caracterizado por sintomas graves localizados no local da picada (WASSERMAN; ANDERSON, 1984). Momentos após a picada, entre 2 e 8 horas, o paciente apresenta dor local, sensação de queimação, vermelhidão e edema, e em alguns casos esses sintomas são acompanhados de prurido (FUTRELL, 1992; MACHIAVELLO, 1947). Esse quadro inicial se deve ao espessamento do endotélio vascular, acúmulo de células inflamatórias, vasodilatação, coagulação intravascular, degeneração dos vasos sanguíneos e hemorragia local para a derme. Também pode ser observada isquemia e posteriormente um quadro de necrose no local da lesão (WASSERMAN; ANDERSON, 1984). Conseqüentemente surge uma mancha arroxeadada no local, a qual se torna ao passar do tempo mais escura chegando à coloração negra com uma região com zonas pálidas e brancas, denominada de placa marmórea (Figura 5). (MACHIAVELLO, 1947; WASSERMAN; ANDERSON, 1984).

O Loxoscelismo sistêmico apresenta sintomas que ocorrem quando a peçonha atravessa a barreira da pele e alcança a circulação sanguínea. Ocasionalmente causa uma degeneração do parênquima visceral, acometendo órgãos internos como fígado e rins, gerando graves processos inflamatórios nos mesmos (MACHIAVELLO, 1947). Os sintomas nesse caso, são febre, mal-estar, prurido, exantema, dor local, calafrios, cefaléia, prostração, náusea e vômitos, icterícia, dispnéia, insônia, taquicardia, hipertermia, hipotensão, perturbações da consciência, coma, choque, e, por fim, congestão e edema pulmonar. Em aspectos biocelulares, podem ser observados coagulação intravascular disseminada, trombocitopenia, hematúria,

hemoglobinúria, cianose, anemia e, em alguns casos restritos, hemólise (MINISTÉRIO DA SAÚDE, 2012; SCHENONE et al., 1989; FUTRELL, 1992). Dados revelam que esse tipo de loxoscelismo é mais comum em regiões onde é encontrada a espécie *L. laeta*, como no estado de Santa Catarina, Brasil (13,1%), no Peru (27,2%), e no Chile (15,7%) (DA SILVA et al. 2004; HOGAN et al. 2004).



Figura 5 Evolução de lesão ocasionada por picada de *Loxosceles*. (A) Local da lesão 48 horas após incidente. O edema característico e áreas de eritema e regiões hemorrágicas (placa marmórea). (B) Outro paciente com uma crosta necrótica 26 dias depois de uma suspeita de picada de aranha. (C) O mesmo paciente com a formação de cicatriz 113 dias após a picada (D) 3 anos após a picada. (Fonte: HOGAN, 2005)

O acidente pode ser classificado segundo o Ministério da Saúde (2012), de acordo com os aspectos clínicos e laboratoriais em leve, moderado e grave. No grau leve, deve ser afastada qualquer hipótese de infecção de pele, dermatite de contato, picadas de insetos ou de outras aranhas. Já no grau moderado, a presença de lesão bem característica com menos de 3 cm é acompanhada de *rash* cutâneo. Nos casos graves, observa-se a presença de lesão característica com mais de três centímetros de diâmetro e/ou alterações clínico-laboratoriais indicativas de hemólise intravascular (MINISTÉRIO DA SAÚDE, 2012).

Atualmente no Brasil, o tratamento adequado nos casos de Loxoscelismo consiste, tanto na soroterapia quanto o tratamento de suporte (MINISTÉRIO DA SAÚDE, 2012) e é específico para cada caso, que incluem de administração de analgésicos e antiinflamatórios a soros anti-

loxoscélico e anti-aracnídico. Na maioria dos casos, a terapia mais indicada é o uso de soro antiloxoscélico concomitante à administração de corticóides (Quadro1) (GUILHERME et al. 2001).

Quadro 1 Classificação quanto à gravidade, quadro clínico e tratamento dos acidentes por aranhas-marrom

*SALox: Soro Antiloxoscélico; **SAAr : Soro Antiaracnídico.

Classificação	Quadro mento	Clínico Trata
Leve	Lesão indefinida, (sendo a <i>Loxosceles</i> identificada como causador do acidente), lesão local sugestiva, sem comprometimento do estado geral e sem alterações após a picada	Sintomáticos: Analgésico, Antiinflamatório local e acompanhamento até 72 hs laboratoriais.
Moderado	Com ou sem identificação da <i>Loxosceles</i> no momento da picada, lesão sugestiva com <i>rash</i> cutâneo ou lesão característica < 3cm de diâmetro, com ou sem alterações sistêmicas e sem alterações laboratoriais, sugestivas de hemólise.	Prednisona (5 dias): Adultos 40 mg/dia e Crianças 1 mg/kg/dia SAAr ou SALox IV 5 ampolas
Grave	Lesão característica de instalação rápida (primeiras 36 h) de diâmetro (forma cutânea) evidência de hemólis, (palidez, anemia aguda, icterícia) confirmada. laboratorialmente (forma cutâneo-visceral)	SAAr ou SALox IV: 5 ampolas a 10 ampolas Prednisona (7 a 10 dias): Adultos 40 mg/dia e Crianças: 1 mg/kg/dia

Fonte: Brasil. Ministério da Saúde. Manual de Diagnóstico e Tratamento de Acidentes por Animais Peçonhentos. 2001.

O Soro Antiloxoscélico, além de diminuir o processo local o tempo de recuperação do paciente, tem como objetivo a neutralização da peçonha circulante e desse modo diminuir a probabilidade da disseminação sistêmica da mesma e impedir complicações que podem levar a óbito, como a hemólise, insuficiência renal e coagulação intravascular disseminada (PAULI et al, 2006). O soro Antiaracnídico, é eficaz contra as peçonhas das aranhas dos gêneros *Loxosceles* e *Phoneutria* e o soro antiloxoscélico é específico sobre a ação da peçonha de *Loxosceles*, sendo indicados quando é identificado o processo de hemólise no quadro cutâneo, caso o diagnóstico seja feito nas primeiras 72 horas (MINISTÉRIO SAÚDE, 2012).

1.3 Epidemiologia do Loxoscelismo

O Loxoscelismo pode ser reconhecido em uma ampla variedade de regiões do mundo (ANDERSON, 1991; BORKAN et al., 1995). No Brasil, o primeiro caso de Loxoscelismo foi diagnosticado no ano de 1954 na cidade de São Paulo, no Hospital Vital Brazil, integrante do Instituto Butantan (CARDOSO & DE CILLO, 1990).

A aranha-marrom, *L. laeta*, foi descrita pela primeira vez em 1947 por Macchiavello, quando foi caracterizada como responsável pela lesão dermonecrótica em um levantamento de 25 casos. Os primeiros casos de uma lesão necrótica registrados, denominada então de “mancha gangrenosa”, de etiologia anteriormente desconhecida em humanos, foram relatados entre os anos de 1800 e 1900, por especialistas sul-americanos (FOIL; NORMENT, 1979).

No Loxoscelismo, os membros inferiores e superiores são as regiões do corpo mais susceptíveis às picadas, já que a maioria dos acidentes ocorrem quando os pacientes estão dentro de suas casas, realizando atividades cotidianas, se vestindo ou dormindo (SEZERINO ET AL. 1998; RIBEIRO et al. 1993).

O Loxoscelismo também representa um problema de saúde pública grave no Peru. Em 2007, Panaftosa relata que o número de acidentes causados por aranhas *Loxosceles* atinge 2500 ao ano. As espécies de maior incidência no Peru são a *L. rufipes* e seguida da *L. laeta*, e os acidentes são mais frequentes em cidades ao longo da costa peruana (GUIMARÃES, et al, 2013; PANAFOTSA, 2007; SANABRIA E ZAVALITA, 1997).

No Brasil, as regiões de maior incidência de casos por acidentes por picadas de aranhas *Loxosceles* são a sudeste e sul, onde os estados mais afetados são Paraná e Santa Catarina, somando 79% dos casos. No geral, os estados mais afetados são Rio Grande do Sul, Santa Catarina, Paraná, São Paulo e Minas Gerais (MARQUES DA SILVA & FISCHER, 2005; MINISTÉRIO DA SAÚDE, 2016).

1.4 A Peçonha da *Loxosceles*

A peçonha loxoscélica trata-se essencialmente de uma mistura protéica de componentes bioquimicamente ativos, cristalina, composta essencialmente por proteínas enzimáticas e não enzimáticas. É produzida por glândulas apócrinas situadas no cefalotórax do animal, a qual demonstra presença abundante de enzimas, diversas toxinas, sendo que a maior parte das proteínas desse composto são de baixa massa molecular (FUTRELL, 1992; GAJARDO e

TOBAR, 1966; DOS SANTOS et al., 2000; da SILVA et al., 2004; da SILVEIRA et al., 2002). De acordo com os respectivos perfis eletroforéticos das proteínas de diferentes famílias das peçonhas loxoscélicas, pode ser encontrado um grupo de peptídeos que se encontra no intervalo de tamanho entre 5- 10 kDa, um grupo formado principalmente por proteínas dermonecróticas com massa molecular encontradas no intervalo de 30-40 kDa, um grupo na qual se encontram as metaloproteases com tem massa molecular mais elevada que os grupos anteriores e finalmente um grupo de proteínas com alta massa molecular situado no intervalo de 60-95 kDa, do qual fazem parte as serino-proteases (CHAVES-MOREIRA, 2011; da SILVA et al., 2004; SILVESTRE et al., 2005).

A peçonha dessas aranhas é inoculada na vítima em volume de poucos microlitros e contém cerca de 30 µg de proteínas. Elas apresentam diferenças importantes relacionadas a expressão e características do composto como um todo, independente da homologia quanto às sequências peptídicas (TAMBOURGI et al, 2010; BARBARO et al., 1996b).

As peçonhas das aranhas peruanas são pouco conhecidas, havendo poucas informações específicas sobre as características estruturais e funcionais dos componentes do arsenal tóxico loxoscélico de *L. laeta* naturais da região (CHÁVEZ-OLORTEGUI, et al, 2013). Para espécies de outras regiões, existem estudos importantes com caracterização da peçonha (TAMBOURGI et al, 2010; OLIVEIRA, 2005; BARBARO et al., 1996a; BARBARO et al., 1996b). As três principais espécies brasileiras de importância médica, *L. intermedia*, *L. gaucho* e *L. laeta*, apresentam atividade dermonecrótica da peçonha semelhantes entre si. Porém, estudos revelam que a peçonha de *L. intermedia* é a mais potente entre elas, com maior letalidade, e o de *L. laeta* é o menos tóxico, mas apresenta um maior potencial dermonecrótico e hemolítico (OLIVEIRA, 2005; BARBARO et al., 1996a; BARBARO et al., 1996b). Barbaro et al, 1996 mostraram que, as peçonhas de *L. laeta* e *L. intermedia* são bem mais complexas que a de *L. gaucho* e que em uma eletroforese SDS-PAGE convencional, o principal componente das peçonhas migra a um padrão de 32 kDa para *L. laeta* e 35 kDa para *L. intermedia* e *L. gaucho*. Entre as várias toxinas descritas na peçonha, as Fosfolipases D ou toxinas dermonecróticas, apresentam um tamanho de 32 kDa e são consideradas as mais expressivas e mais bem caracterizadas enzimas presentes na peçonha da aranha-marrom.

A característica mais marcante da peçonha da aranha-marrom é determinada por seus efeitos dermonecróticos (BARBARO et al., 1992). Segundo estudos anteriores, essas enzimas correspondem ao componente responsável pela alta toxicidade da peçonha, já que a agregação plaquetária, a hemólise, o edema, a lesão dermonecrótica e o aumento da permeabilidade vascular foram relacionados à atividade dessa enzima (APPEL et al., 2006; FUTREL, 1982; da

SILVA et al., 2004; PALUDO et al., 2009). Além dos sintomas característicos, alguns efeitos sistêmicos são bem peculiares no desenvolvimento da patologia, como a coagulação intravascular disseminada e a insuficiência renal aguda (MACHIAVELLO, 1947A; REES et al., 1984; WILLIAMS et al., 1995). No local da lesão, supõe-se que a ativação do complemento e a infiltração de Polimorfonucleares (PMN) contribuem para o desenvolvimento desse quadro (Futrell, 1992). A ativação do sistema complemento é feita através de uma via alternativa em que acontece a deposição de C3 e Fator B que acarretará na organização da C3 convertase e do complexo responsável por danificar a membrana celular (TAMBOURGI et al., 2000). Um estudo realizado por Pizzi et al. (1957) em cobaias, revelou um quadro de vasoconstrição inicial das arteríolas anterior a um processo de vasodilatação intensa. O tecido dos vasos sanguíneos dos animais adjacentes à área da lesão apresentou um processo de degeneração e trombose. Nas primeiras horas há formação de um intenso edema com extravasamento de células sanguíneas, inclusive eritrócitos. Outros estudos mostraram que células inflamatórias, especialmente polimorfonucleares, são mais concentradas no local da lesão, juntamente com a infiltração leucocitária que gera injúria nos vasos sanguíneos o que ocasiona um processo de perda da derme (PATEL et al., 1994; SMITH & MICKS, 1970).

Algumas isoformas diferentes de Fosfolipases D já foram identificadas e foram denominadas como proteínas Loxtox (KALAPOTHAKIS et al., 2007). Elas descritas em várias espécies de *Loxosceles*, demonstrando a existência de várias formas ativas e inativas de fosfolipases D em uma mesma peçonha (FERNANDES-PEDROSA et al., 2002; KALAPOTHAKIS et al., 2007; MACHADO et al., 2005). Na peçonha de *L. reclusa*, existem 4 isoformas de fosfolipases D já descritas (RIBEIRO et al., 2007). Essas diferenças podem se relacionar a uma adaptação durante a evolução das espécies de *Loxosceles* quando submetidas a ambientes diversos para sua sobrevivência (MACHADO et al., 2005).

Em estudo anterior, Tambourgi et al., 2000 e Gremski et al (Fig.6) relacionaram também outras moléculas, de 33, 34 e 35kDa aos efeitos nocivos da peçonha da aranha-marrom. Os autores realizaram uma classificação funcional dos transcritos, através da técnica de Sanger de sequenciamento para identificação de ESTs, da glândula da peçonha de *L. laeta*, mostrando a proporção relativa de diferentes tipos de transcritos: toxinas conhecidas (16%), possíveis toxinas (14,5%) e proteínas celulares (33,3%), e sequências sem correspondências (24,6%).

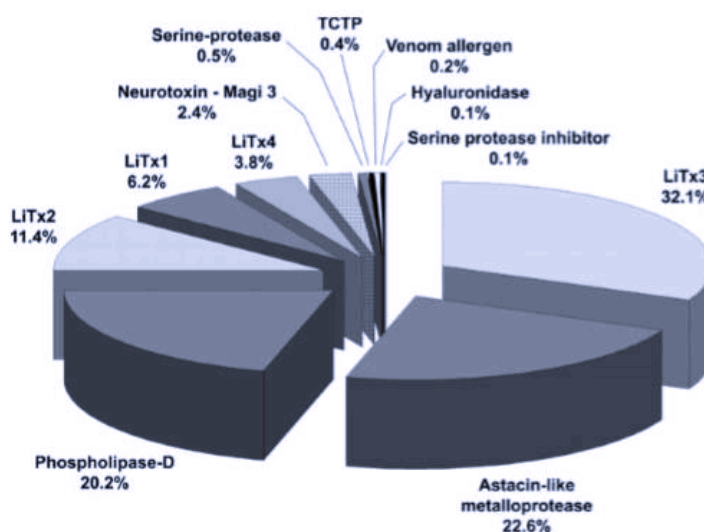


Figura 6 Distribuição da porcentagem do nível de expressão de cada grupo de toxinas e enzimas produzidas na glândula de peçonha da aranha *L. intermedia*. Fonte: GREMSKI et al., 2010.

Em menor concentração que as Fosfolipases, Gremski et al (2010), identificaram na glândula da peçonha de *L. intermedia* muitos transcritos que codificam ICKs. Também já foram encontradas no arsenal da peçonha as metaloproteases da família das astacinas (LALP-Loxosceles Astacin-like Protease) - Metaloproteinases com 20-28 kDa, responsáveis pelos efeitos fibronectinolíticos e fibrinogenolíticos; e as Metaloproteinases gelatinolíticas de 32-35 kDa. Alguns estudos relacionam essas enzimas a atividade de digestão da presa da aranha, bem como hemorragias e no auxílio na dispersão de outras toxinas no corpo da presa (da SILVEIRA et al., 2007b, TREVISAN-SILVA et al., 2010). O mecanismo de ação dessas enzimas consiste na degeneração da laminina, colágeno e fibronectina da matriz extracelular (KLEINER AND STETLER-STEVENSON, 1993).

Alguns estudos já identificaram a presença de serino-proteases na peçonha de aranha-marrom (VEIGA et al., 2000) e inibidores de serino-proteinases, que podem representar uma papel de proteção para as toxinas da peçonha (ZUPUNSKI et al., 2003; FERNANDES-PEDROSA et al., 2008; MACHADO et al., 2005).

Outros estudos mostram a existência de outros componentes importantes da peçonha loxoscélica, as quais são as hialuronidasas, TCTP (Translationally Controlled Tumor Protein) e peptídeos inseticidas (CHAIM et al., 2006; DE CASTRO et al., 2004; FEITOSA et al., 1998; KUSMA et al., 2008; MACHADO et al., 2005; RIBEIRO et al., 2007; TAMBOURGI et al., 2000; VEIGA et al., 1999; VEIGA et al., 2000b). As hialuronidasas degradam o ácido hialurônico que consiste em um componente encontrado na matriz extracelular, ocupando

espaços intercelular e cumprindo papel de barreira protetora contra o acesso de partículas externas (MARKOVIC-HOUSLEY et al., 2000). Estudos mostraram a relação dessa enzima com a disseminação da peçonha pelo organismo e potencialização dos efeitos dermonecroticos (FERRER et al., 2013).

1.5 A Bioinformática na Transcriptômica

1.5.1 Introdução ao estudo computacional de transcriptomas

O transcriptoma é o conjunto completo de transcritos a partir de um determinado genoma e sua quantificação em um determinado tempo, estágio de desenvolvimento específico ou condição fisiológica (WANG et al., 2009). Ferramentas computacionais para a análise desses dados de biologia molecular são cada vez mais essenciais devido ao grande número de informações a serem processadas. Uma grande variedade de sequências é gerada a cada pesquisa realizada, o que está relacionado ao aperfeiçoamento da metodologia de sequenciamento (MARDIS, 2011). A análise de todo este conjunto de dados soma um grande número de informações para o enriquecimento dos dados genômicos e transcriptômicos de diferentes espécies, fazendo possível a construção de bibliotecas para os mais diversos estudos (AUBOURG E ROUZÉ, 2001). A bioinformática revelou-se essencial para o estudo desta complexidade biológica, que está relacionada com o desenvolvimento e execução de métodos de armazenamento, análise, modelagem e descoberta de conhecimento a partir de dados biomédicos (KASABOV, 2007).

No presente estudo, foi realizada uma análise da diversidade molecular de fosfolipases D da peçonha da aranha *L. laeta* peruana utilizando tecnologia de nova geração ou *Next Generation Sequencing* (NGS). Essa tecnologia proporciona um número significativamente maior de transcritos, que podem incluir os inéditos e os pouco expressos, em comparação com a técnica de sequenciamento capilar de Sanger. Inicialmente, na década de 70, a equipe de Frederick Sanger desenvolveu o método dideoxynucleotídeo ou ddNTP, também conhecido como método Sanger de sequenciamento (SANGER; et al., 1977). Essa técnica, considerada na época padrão ouro para o sequenciamento de ácidos nucleicos, foi a mesma utilizada no Projeto Genoma Humano para determinação das informações genéticas de nossa espécie (GRADA; WEINBRECHT, 2013). Muitos transcriptomas de diferentes espécies foram caracterizados com base em ESTs sequenciados com a tecnologia capilar. Contudo, trata-se de uma tecnologia

de custo elevado, com muitas etapas experimentais de manipulação trabalhosa, que exige demanda elevada de tempo (STRAUSBERG et al., 2008).

Ao longo dos anos seguintes, houve um aumento considerável da demanda por métodos de sequenciamento mais eficientes com menos erros que necessitem de menos tempo de execução e gerem maior quantidade de dados com a mesma quantidade de recursos (ANSORGE, 2009; GRABHERR *et al.*, 2011). Com a evolução dessas tecnologias, foi possível expandir as análises e identificação de RNAs expressos em um indivíduo ou tecido, bem como a comparação do perfil de expressão gênica das amostras em distintas condições ambientais, contexto fisiológicos ou estágios de desenvolvimento, sendo possível o estudo de genes anteriormente anotados ou nunca antes estudados (ANSORGE, 2009; GRABHERR et al., 2011). A evolução do sequenciamento de Sanger para o NGS, possibilitou a determinação do aumento da padrão da montagem, levando em consideração as maiores possibilidades para controle da qualidade das *reads* brutas (BONFIELD et al., 1995). Técnicas NGS acarretaram em uma importante revolução no campo científico e proporcionaram pesquisas biológicas em um nível de grande escala. Sendo essas técnicas (NGS) mais baratas e otimizadas, é possível sequenciar um genoma inteiro, transcriptoma ou exoma, com baixo índice de erros demandando um intervalo de tempo interessante e um menor custo. (PAVLOPOULOS et al., 2013).

Com o NGS foi possível fazer uso de nanotecnologias que possibilitaram o sequenciamento em larga escala de fragmentos pequenos de DNA e cDNA e a obtenção de Gigabases (Gb) de seqüências genoma ou transcriptoma de qualquer espécie (PAVLOPOULOS et al., 2013). Devido à expansão das aplicações das tecnologias de sequenciamento de alto rendimento e da acessibilidade a essas metodologias, ferramentas computacionais para a análise desses dados, são cada vez mais essenciais (WANG et al., 2009). Isso torna favorável a caracterização de polimorfismos associados aos genes transcritos: *splicing* alternativo e SNPs; a compreensão dos padrões de metilação do DNA e das modificações no posicionamento das histonas e a análise dos perfis de expressão de *micro* RNAs e das interações entre proteínas e entre DNA e proteínas (LISTER et al., 2009; WANG et al., 2009; GROBA & BURGOS, 2010; GARBER et al., 2011).

Dentre as principais plataformas de sequenciamento NGS está a Illumina, que foi utilizada nesse trabalho. Basicamente, as etapas do processo de sequenciamento do Illumina consistem em preparar as amostras de transcriptoma ou genoma a serem analisadas através da ligação de adaptadores nas extremidades 5' e 3' dos fragmentos do material. Sequências complementares a eles se ligam a suas seqüências quando estes são dispensados sobre uma superfície de sequenciamento sólida preenchida de oligonucleotídeos (SHENDURE, 2008). Os

fragmentos aderidos à placa passam por um processo de clonagem por PCR em “fase sólida”. O oligonucleotídeos que haviam sido aderidos na superfície da placa podem ser chamados de *primers* nesse contexto. São geradas muitas réplicas de um mesmo fragmento oriundo da amostra original. (MARDIS, 2008; METZKER, 2010).

A plataforma Illumina, além de apresentar uma excelente performance e otimizar o processo de obtenção de resultados, apresenta custo reduzido e aumenta a profundidade do sequenciamento por amostra. Assim como as outras plataformas de NGS, a Illumina produz elevado número de sequências, porém, seus *reads* relativamente curtos representavam limitações para o estudo de alguns tipos de amostras, sendo solucionado com o surgimento da plataforma MiSeq-Illumina com *reads* pareados de 250 a 300 pb gerando resultados mais precisos e confiáveis (WANG et al., 2007; JEON et al., 2015).

1.5.2 RNA-Seq e a montagem *de novo*

O RNA-Seq é uma metodologia de NGS, que revela o sequenciamento de DNA complementar (cDNA), apresenta alta sensibilidade e pode ser utilizada para caracterizar o transcriptoma e o genoma de um organismo (PINTO et al., 2011). Isto oferece diversas vantagens sobre as tecnologias existentes (Tabela 1). Além de encontrar transcritos que correspondem à sequência original de uma amostra com genes ou transcritos, também pode encontrar variações de sequência (WANG et al., 2009).

Através do RNA-seq um conjunto sequências de RNA é convertido em uma biblioteca de cDNA, e gera milhões de fragmentos de sequências curtas (*reads*) (Figura 7). Para calcular níveis de expressão e identificar *splicing* alternativo, estes fragmentos são alinhados contra sequências de referência ou as próprias sequências obtidas (TRAPNELL et al., 2009). Essa técnica é utilizada para a análise quantitativa de níveis de expressão de transcritos, gerando dados que fornecem medições de níveis de transcritos em variadas condições (JÄNES et al., 2015).

Tabela 1. Vantagens do método RNA-Seq comparado aos outros métodos de sequenciamento.

Tecnologia	Sequenciamento de <i>Sanger</i>	<i>Microarray</i>	RNA-Seq
Princípio	Sequenciamento de <i>cDNA</i> e ESTs	Hibridização	Sequenciamento em larga escala
Resolução	Única base	Variação até 100 pb	Única base
Dependência de genoma de referência	Não	Sim	Algumas vezes dispensável

Ruído	Baixo	Alto	Baixo
Throughput	Baixo	Alto	Alto
Aplicações			
Mapear regiões transcritas e expressão gênica	Limitado	Sim	Sim
Quantificação do nível de expressão gênica	Limitado	Variação até 100	> 8000 - vezes
Distinção de diferentes isoformas	Sim	Limitado	Sim
Distinção da expressão de alelos	Sim	Limitado	Sim
Demanda prática para execução			
Quantidade de amostra	Alta	Alta	Baixo
Custo	Alta	Alta	Baixo

Fonte: Adaptado de WANG et al., 2009.

O processo de montagem das *reads* obtidas é realizado através de ferramentas de bioinformática específicas para cada tipo de metodologia de análise. Dentre as diversas ferramentas desenvolvidas para a análise de dados transcriptômicos, as mais citadas na literatura para montagem quando há uma sequência de referência são o TopHat (GHOSH e CHAN, 2016), o Cufflink (GHOSH e CHAN, 2016) e o Trinity (GRABHERR et al., 2011) (GARBER et al., 2011 e TRAPNELL et al., 2012).

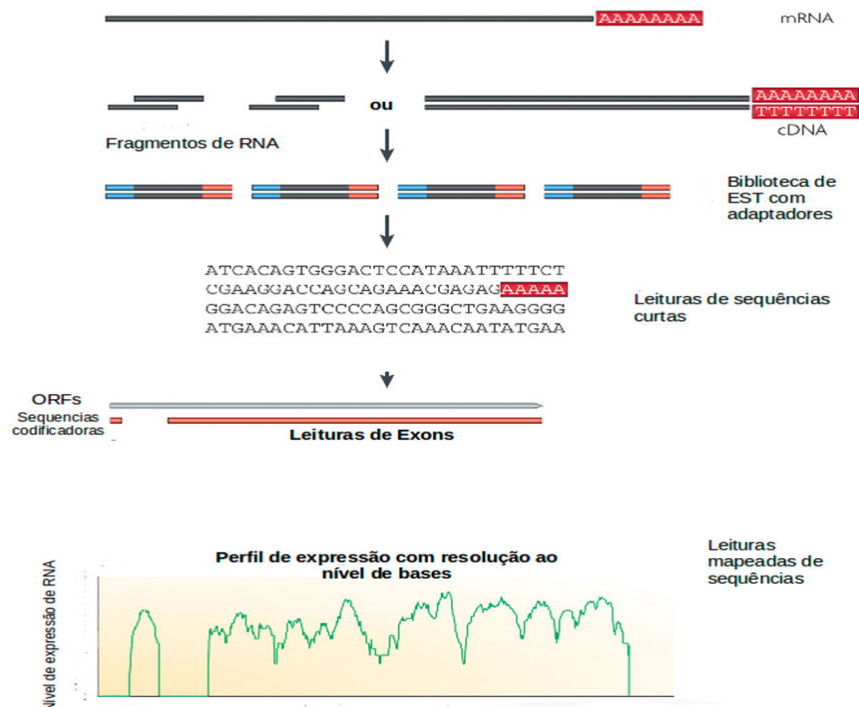


Figura 7 Esquema mostrando metodologia de RNA-Seq. O RNA é fragmentado previamente à montagem da biblioteca de cDNA. As seqüências obtidas são alinhadas a um

genoma de referência para a montagem dos *contigs* e quantificação da expressão. Fonte: modificado de WANG et al, 2009

Para a montagem de novo, o Velvet-Oases (SCHULZ et al., 2012) e o Trinity (GRABHERR et al., 2011) são programas muito utilizados. Os algoritmos implementados nesses programas, bem como outros criados para a montagem de novo de transcriptomas, foram analisados e comparados em diversos trabalhos recentes, sendo que o software Trinity é citado como uma das ferramentas mais eficientes na montagem de novo de transcriptomas, sendo considerada padrão ouro na bioinformática (GROBA e BURGOS 2010; HASS et al., 2013). Esse programa é baseado no gráfico de Bruijn (NAGARAJAN and POP, 2013). Este algoritmo utiliza abordagens em que as *reads* são divididas em k-mers, de tamanho k, onde cada aresta do gráfico correspondente a um k-mer da sequência (PEVZNER; TANG; WATERMAN, 2001), esquema que será melhor explicitado da metodologia.

2 JUSTIFICATIVA

O Loxoscelismo é um problema de saúde pública tanto no Brasil quanto no Peru. Agravado pelo crescente número de casos de acidentes em ambos os países, pelos hábitos cosmopolitas adquiridos pelas aranhas do gênero *Loxosceles* e pela complexidade da peçonha desses indivíduos. A identificação dos compostos da peçonha favorece a melhor compreensão dos processos fisiopatológicos do envenenamento, bem como o desenvolvimento de ferramentas para melhorias nos métodos de profilaxia e tratamento para o Loxoscelismo. Devido à grande complexidade da peçonha da aranha *Loxosceles*, nesta dissertação focamos nossos estudos transcriptômicos nas proteínas da peçonha responsáveis pela principal atividade tóxica do composto, que são as Fosfolipases D, esfingomielinases D ou proteínas dermonecróticas. Além das vantagens referentes ao loxoscelismo, o estudo aprofundado da expressão dessas enzimas na glândula de peçonha e sua constituição, favorecem estudos sobre a filogenia dessas proteínas e contribui para a taxonomia das próprias espécies do gênero.

3 OBJETIVOS

3.1 Objetivo Geral

Identificar e descrever as Fosfolipases D da glândula de peçonha de aranhas da espécie *Loxosceles laeta*, provenientes do Peru, através de sequenciamento de nova geração e análise transcriptômica.

3.2 Objetivos Específicos

- Construir uma biblioteca de cDNA das glândulas de peçonha da aranha *Loxosceles laeta*;
- Sequenciar a biblioteca gerada utilizando tecnologia de nova geração;
- Realizar a montagem de novo das *reads* geradas;
- Identificar os *contigs* contendo sequências similares às de fosfolipases-D e fazer a anotação das sequências;
- Analisar o nível de expressão relativo entre todos os *contigs* alinhados às fosfolipases D;
- Classificar a atividade das proteínas identificadas com base na relação filogenética de sua sequência de aminoácidos com outros peptídeos de função conhecida.

4 METODOLOGIA

4.1 Extração de peçonha de *L. laeta*

A peçonha foi extraída no Laboratório de Biotecnologia e Marcadores Moleculares (ICB-UFMG), através da dilaceração da glândula de peçonha de cerca de 5 aranhas da espécie *Loxosceles laeta* fêmeas coletadas no Peru. As glândulas de peçonha dos indivíduos foram removidas, o material foi homogeneizado em 500µL de TRIZOL por 50 a 100 mg de tecido, e incubados por 5 a 10' a uma temperatura de 15° a 30° C, para permitir a completa dissociação de complexos nucleoprotéicos. Logo em seguida, foi realizada a maceração com auxílio de um triturador mecânico (IKA 710 Basic Ultra-turrax).

A solução obtida foi mantida em repouso por 10 minutos e em seguida acrescentado 100 µL de clorofórmio (Synth). A amostra foi centrifugada a 12.000 x g por 15' a 4° C e em seguida separado o RNA em uma fase aquosa e feita a remoção do sobrenadante. O RNA foi precipitado da fase aquosa pela adição de isopropanol, homogeneizado e incubado por 10' a 15 a 30°C.

Feito isso foi realizada uma nova centrifugação a 12.000 x g por 10' a 4°C com o fim de precipitar o RNA. O sobrenadante no meio aquoso foi recolhido e aplicado em um Falcon, em seguida o RNA foi precipitado com 300 µL de isopropanol (Synth) e incubado por 15 minutos a temperatura ambiente.

O precipitado de RNA frequentemente invisível nas centrifugações anteriores, foi então separado no vortex e novamente centrifugado a 12.000 x g por 10' a 4° C. O precipitado foi lavado com 500 µL de etanol 80% por duas vezes e o RNA foi seco à temperatura ambiente. Posteriormente, o RNA foi ressuscitado com 100 µL água ultrapura (Phoneutria) e incubado por 30 minutos a 56°C para a completa dissolução do precipitado.

Completa essa etapa, foi realizada a quantificação em NanoDrop e análise de qualidade em eletroforese em gel de agarose 1% feito com água DEPC e TAE DEPC e estocadas as amostras a -80° C.

4.2 Construção da biblioteca para sequenciamento de nova geração

A biblioteca foi construída utilizando o kit TruSeq™ RNA Sample Prep Kit v3 Set A (Illumina), segundo o protocolo fornecido pelo fabricante. A construção da biblioteca inicia com a purificação do mRNA utilizando *beads* magnéticas contendo oligo (dT). Em seguida, o mRNA purificado foi fragmentado e feita a síntese do cDNA (1ª e 2ª fita) com a enzima Superscript II e as extremidades das fitas foram reparadas para a geração de extremidades cegas. A seguir, foi realizada a ligação dos adaptadores com índice (sequência código para identificação de amostras). A quantificação foi feita com o kit de PCR quantitativa “Library Quantification Kit - Illumina/Universal” (Kapa), de acordo com as instruções do fabricante. A biblioteca quantificada foi diluída e aplicada no cartucho de sequenciamento após denaturação.

4.3 Sequenciamento de Nova Geração (NGS) – MiSeq Illumina

As bibliotecas foram sequenciadas através do sequenciador de nova geração MiSeq (Illumina), pela equipe do Prof. Dr. Evanguedes Kalapothakis, no Laboratório de Biotecnologia e Marcadores Moleculares do ICB/UFMG. Foram utilizados dois cartuchos de sequenciamento, sendo que a execução do processo gerou 600 ciclos dos quais 1 com 300 ciclos para as reads 3'5', e outro de 300 ciclos para as reads 5'3', com a estratégia de sequenciamento *paired-end*.

As bibliotecas quantificadas foram diluídas em uma concentração de 4 nM. Em seguida, foi realizado um pool com 10 μL de cada biblioteca e o volume final homogeneizado com auxílio de um vortex. A desnaturação do pool de bibliotecas (5 μL) foi realizada por 5 minutos com 5 μL de NaOH 0,2 N. Após a incubação para desnaturação, foi adicionado às bibliotecas desnaturadas 990 μL de solução HT1 do kit de sequenciamento. As bibliotecas desnaturadas obtidas na concentração de 20 pM foram novamente diluídas para a concentração de 15 pM com a adição de 250 μL de solução HT1 e 750 μL da biblioteca. Por fim, as bibliotecas desnaturadas em pool na concentração de 15 pM (600 μL) foram aplicadas nos cartuchos de sequenciamentos.

O resultado do sequenciamento gerou um arquivo no formato fastq, típico dessa metodologia de sequenciamento, o qual contém as bases e seus valores, já com sequências adaptadoras removidas pelo sequenciador MiSeq, originando uma biblioteca com as *reads* brutas.

4.4 Pré-processamento das *reads*

O arquivo gerado a partir do sequenciamento da amostra passou subsequentemente por um processo de trimagem das *reads*. A trimagem das sequências consiste na remoção de trechos pré-definidos, com baixa qualidade de bases e adaptadores, que possam interferir na futura montagem e anotação dos dados (GUO et al., 2013). Para obter um bom resultado dessas trimagens, inicialmente é preciso realizar uma análise da qualidade das *reads* (DAI et al., 2010; PABINGER et al., 2014).

Os arquivos de saída do sequenciador, em formato fastq, apresentam um valor de qualidade (Q), *Phred quality score*, para cada base sequenciada. Esses valores têm relação logarítmica com a probabilidade de erro (P) de nomeação da base, de acordo com a fórmula $Q = -10 \log P$. Assim, para um valor de $Q = 20$, temos a probabilidade de 1 base errada em 100 bases nomeadas, como representado na Tabela 2 (Ewing et al., 1998). Nesse trabalho, todas as bases com valor menor do que Q20 foram retiradas das *reads* a partir da extremidade 3', assim como as *reads* com tamanho inferior a 40 pb (EWING; GREEN, 1998; EWING et al., 1998)

Tabela 2 Valores de Phred e correspondente probabilidade de erro de nomeação das bases. O valor de Phred (Q) = $-10 \log_{10} P$,

onde a probabilidade de erro de nomeação de base é representada por (P).

Valores de <i>Phred</i>	Probabilidade de nomeação incorreta de bases (P)	Porcentagem de precisão de nomeação das bases
10	1 em 10	90%
20	1 em 100	99%
30	1 em 1000	99,90%
40	1 em cada 10.000	99,99%
50	1 em 100.000	100,00%

Fonte: baseada em EWING et al., 1998.

Para a remoção dessas bases de baixa qualidade foi utilizado o programa PRINSEQ lite 0.15 (<http://prinseq.sourceforge.net>) (SCHMIEDER e EDWARDS, 2011). Esse programa, na sua versão *standalone*, utiliza da linguagem Perl (<http://perldoc.perl.org>). A trimagem foi realizada com base na linha de comando abaixo. O significado de cada opção utilizada está especificado no Quadro 2.

```
perl prinseq-lite.pl -verbose -fastq input.fastq -out_good
outputgood -out_bad outputbad -min_len 40 -trim_qual_right (22-
32) -trim_qual_window (50-100) -trim_qual_step 1 -trim_qual_type
mean -out_format 5
```

Quadro 2 Opções do programa Prinseq-lite utilizados para trimagem das sequências de baixa qualidade e suas respectivas funções de acordo com o manual do usuário.

Opção para prinseq-lite	Função do comando e argumento
-verbose	Imprime no terminal o relatório do andamento do processo
-fastq	Formato do arquivo de entrada
-out_good:	Pasta/nome do arquivo com as sequências filtradas
-out_bad:	Pasta/nome do arquivo com as sequências que não passaram pelo filtro
-min_len:	Tamanho mínimo de 40 bases nas reads filtradas
-trim_qual_right:	Retirada das sequências a partir da extremidade 3'
-trim_qual_window:	Trimagem por uma janela de 50 a 100 nucleotídeos

-trim_qual_step:	Tamanho de deslocamento da janela
-trim_qual_type:	Tipo de parâmetro de qualidade a ser usado na trimagem
-out_format:	Formato dos arquivos de saída (fasta, qual e fastq)

Os parâmetros de trimagem foram estabelecidos de forma a obter o maior número de reads com elevados valores de qualidade. Antes e após a trimagem, o programa Fastqc (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>) foi utilizado para obtenção de gráficos para visualização do resultado e direcionar a escolha dos parâmetros até a obtenção do resultado desejado (GUO et al., 2013).

4.5 Montagem dos *contigs*

Após a trimagem, as *reads* resultantes passaram pelo processo de montagem sem que fosse utilizada um genoma de referência (montagem *de novo*) (GRABHERR et al., 2011). A montagem foi realizada por meio da execução do programa Trinity (trinityrnaseq-2.1.1) (<https://github.com/trinityrnaseq/trinityrnaseq.wiki.git>), através do servidor de alta performance, Sagarana, do departamento de Biodados da Universidade Federal de Minas Gerais.

O Trinity é baseado em três algoritmos conforme abaixo (Figura 8):

- Inchworm: reúne o conjunto de dados original em sequências únicas de transcritos;
- Chrysalis: forma clusters a partir dos contigs construídos pelo primeiro algoritmo e constrói gráficos *de Bruijn* para cada cluster, particionando a leitura inicial de Inchworm em gráficos distintos;
- Butterfly: analisa os gráficos individualmente e de acordo com a posição das *reads* e dos pares de reads no gráfico e reporta os transcritos resultantes de *splicing* alternativo bem como os transcritos originados de genes parálogos.

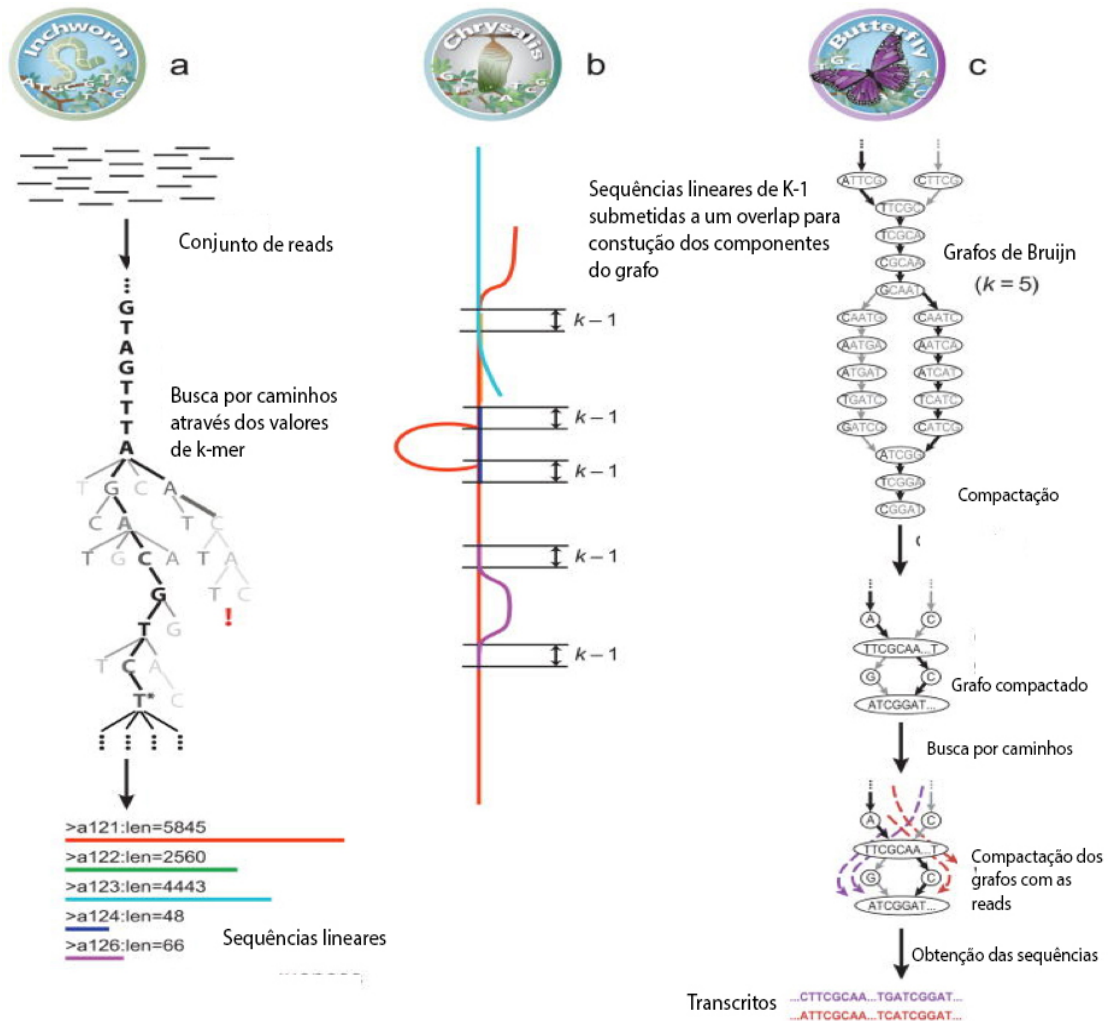


Figura 8 Visão geral do Trinity: (A) Inchworm monta o conjunto de dados de leitura (linha preta curta, parte superior) à procura de caminhos através de valores de K-mer, resultando numa coleção de seqüências contíguas lineares com cada k-mer. (B) Chrysalis reúne contigs que compartilham pelo menos um k-1-mer e faz a leitura da extensão do arquivo a juntar-se, e constrói gráficos individuais de Bruijn de cada grupo (linhas coloridas). (C) Butterfly: relaciona cada de gráfico de Bruijn e de Chrysalis (em cima). Fonte: Haas BJ et al, 2013

Foram realizadas diferentes montagens do mesmo conjunto de dados, aplicando diferentes valores de K-mer (18 a 32). O objetivo era identificar o valor de K mais adequado à futura montagem dos transcritos relacionados para melhores resultados quando alinhados a fosfolipases. O parâmetro padrão utilizado pelo Trinity requer cobertura mínima de 3x e utiliza K igual a 25. A linha de comando utilizada está descrita abaixo e a função de cada opção está especificada no Quadro 3.

```
./Trinity --seqType fq --max_memory 50G --left R1.fastq --right R2.fastq --KMER_SIZE (18-32) --CPU 64 --SS_lib_type FR --output trinity_output
```

Quadro 3 Opções do programa Trinity, e suas funções correspondentes, utilizadas para a montagem de novo das reads.

Opções do programa Trinity utilizadas	Função
--seqType	Formato do arquivo de entrada
--max_memory	Quantidade máxima de memória RAM a ser utilizada na montagem.
--left	R1 no caso de <i>reads</i> pareadas
--right	R2 no caso de <i>reads</i> pareadas
--KMER_SIZE	Tamanho do K-mer a ser utilizado
--CPU	Número de núcleos do processador a serem utilizados
--SS_lib_type	No caso de <i>reads</i> pareadas, fixa que em cada par, as <i>reads</i> pertencem a fitas opostas
--output	Local/nome do arquivo de saída do programa.

Através do software CLC Genomics Workbench 9 (<https://www.qiagenbioinformatics.com>), as *reads* processadas foram analisadas e foi possível dessa maneira observar, através de gráficos, a posição das *reads* em cada um dos *contigs* e assim determinar visualmente a cobertura da montagem de cada sequência consenso. Utilizando a ferramenta RNA-Seq Analysis (Length fraction = 0,9 e Similarity fraction = 0,9) foi calculada a expressão de cada transcrito, definida em RPKM (Reads Per Kilobase of exon model per Million mapped reads). Como *reads* pareadas foram utilizadas no mapeamento e contadas conjuntamente, RPKM equivale a *reads* por kilobase de exon/transcrito por milhão de fragmentos mapeados (http://www.clcsupport.com/clcgenomicsworkbench/current/index.php?manual=Definition_RPKM.html#sec:RPKM).

4.6 Anotação das sequências relacionadas a fosfolipases

Os dados obtidos em formato fasta através da montagem foram submetidos a um alinhamento de sequências com a ferramenta BLASTx *stand-alone*, a fim de conseguir a identificação dos *contigs* com similaridade a fosfolipases. O banco de dados de fosfolipases

utilizado para a realização do Blastx foi obtido pelo site do NCBI (<http://www.ncbi.nlm.nih.gov>). Para essa busca foram utilizados os termos “phospholipase D”, “sphingomyelinase D” e “dermonecrotic protein” com objetivo de filtrar as sequências relacionadas às fosfolipases.

Foi realizada uma filtragem dos alinhamentos obtidos, a qual objetivou descartar os resultados com valores pouco relevantes e os alinhamentos menos significativos para o mesmo contig. Os parâmetros levados em consideração para inferir a relevância foram os seguintes:

- Tamanho do alinhamento acima de 200 aminoácidos (o que corresponde ao tamanho aproximado das phospholipases previamente anotadas na literatura).
- Identidade igual ou maior a 50%
- Eliminação de contigs em repetição, que alinharam com mais de uma sequência referência
- Valor de e-value máximo de 10^{-5} .

As sequências com melhores valores baseados nos parâmetros considerados acima foram selecionadas, ou seja, as sequências que apresentaram maior similaridade com fosfolipases pelo programa Blastx *stand-alone*. Esses *contigs* selecionados foram submetidos à busca de similaridade na plataforma *on line* do programa Blastx.

Com o objetivo de identificar os domínios, peptídeo sinal e região codificadora de cada fosfolipase D, as sequências de cDNA dos contigs selecionados foram traduzidas através na ferramenta *on line* de tradução de sequências, Expasy - SIB (Bioinformatics Resource Portal) (<http://web.expasy.org/translate/>). Em seguida, foi feita uma análise *in silico* de algumas propriedades físico-químicas das proteínas (p.I.) utilizando para isso outra ferramenta disponível através do Expasy (<http://www.expasy.ch/tools/protparam.html>). Essas sequências foram analisadas quanto a presença de peptídeo sinal com auxílio do programa online SignalP 4.0 (PETERSEN et al., 2011).

O Cladograma (WATERHOUSE et al. 2009) foi construído com base em um alinhamento das sequências peptídicas no programa clustal W (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Esse alinhamento proporcionou a confecção de uma figura obtida pelo programa Jalview (www.jalview.org) para estudos didáticos e a criação de um cladograma através do programa Mega 7 (<http://www.megasoftware.net>). O mega 7 disponibiliza uma ferramenta para construção da árvore filogenética que consiste na execução do algoritmo de *neighbor-joining*.

5 RESULTADOS E DISCUSSÃO

5.1 Construção da biblioteca de cDNA

Como era esperado para sequenciamentos Illumina, a avaliação inicial da qualidade do sequenciamento, observada na Figura 9, mostrou a necessidade de trimagem das extremidades 3' das *reads* devido à queda da qualidade das *reads* na porção final. Os arquivos fastq obtidos pelo sequenciamento foram denominados L-laeta_S1_L001_R1_001.fastq e L-laeta_S1_L001_R2_001.fastq, sendo R2 o arquivo resultante do sequenciamento reverso complementar (*Paired-End*). Foram obtidas 11.870.865 *reads*, com comprimento de sequência variável entre 35-301 bases, totalizando 2.392.153.725 bases e conteúdo de GC de 41%.

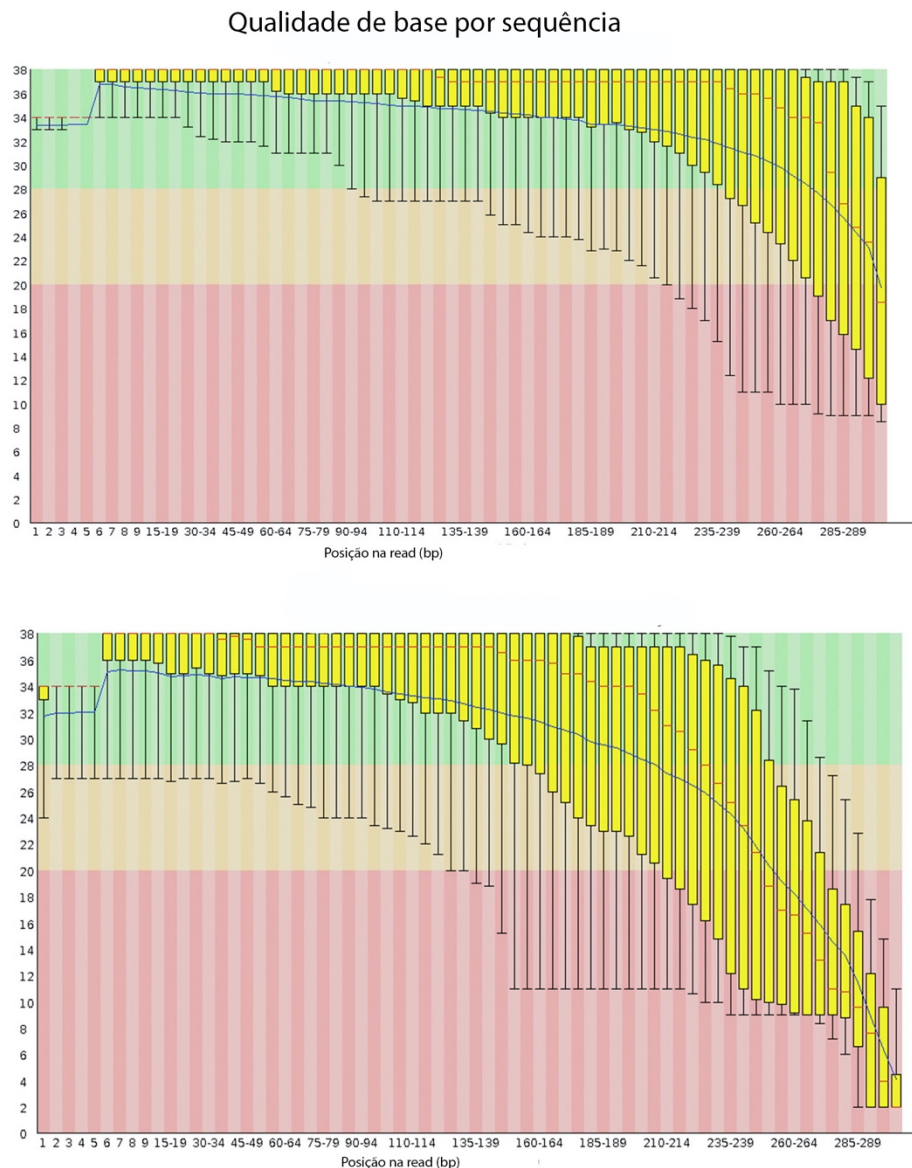


Figura 9 Valores de qualidade Q por base ao longo das *reads* antes da trimagem. A qualidade das *reads* brutas, obtidas após o sequenciamento, foi analisada pelo programa

FastQC. Nos gráficos estão representados os valores médio de qualidade das bases ao longo das *reads*. O eixo y mostra os índices de qualidade. O eixo x representa cada base. As colunas amarelas verticais contém o intervalo interquartil dos valores de qualidade para cada base, mensuradas entre 25% e 75%, onde também pode-se ver linhas em preto acima e abaixo dessas colunas que correspondem aos valores 10% e 90% dos dados. Dispostas dentro de cada barra amarela há uma linha em azul, que consiste na flutuação do valor médio da qualidade e uma linha em vermelho, determinando o valor mediano de qualidade de cada base. **A)** Análise de qualidade realizada das *reads* do arquivo inicial antes da trimagem, para L-laeta_S1_L001_R1_001.fastq. **B)** Análise de qualidade realizada das *reads* do arquivo inicial antes da trimagem, para L-laeta_S1_L001_R2_001.fastq.

5.2 Pré-processamento das sequências

A avaliação inicial da qualidade do sequenciamento, observada na Figura 9, mostrou a necessidade do tratamento nos arquivos de *reads*, que foram trimados separadamente. Como esperado no sequenciamento Illumina, as últimas bases apresentam uma flutuação abaixo de 20. O resultado da trimagem pode ser observado na Figura 10.

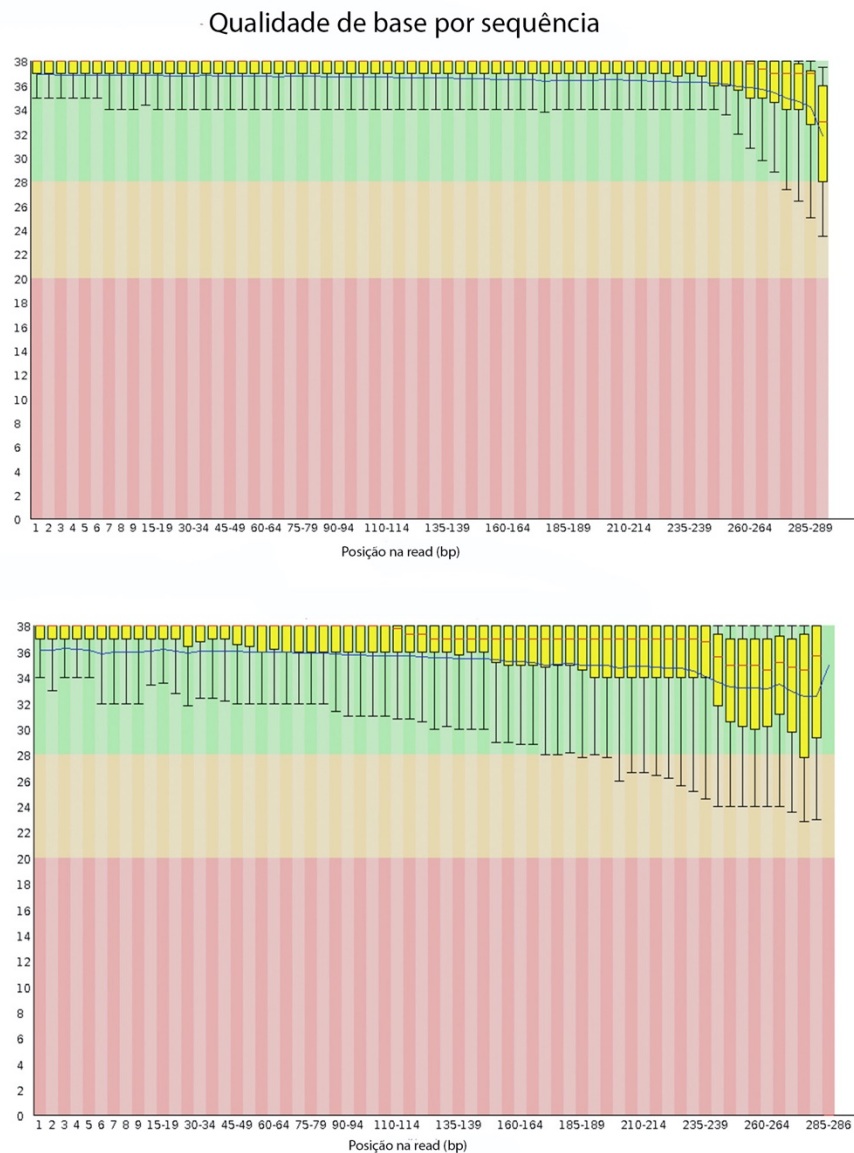


Figura 10 Valores de qualidade Q por base ao longo das *reads* após trimagem. A qualidade das *reads* foi analisada pelo programa FastQC. Nos gráficos estão representados os valores médio de qualidade das bases ao longo das *reads*. O eixo y mostra os índices de qualidade. O eixo x representa cada base. As colunas amarelas verticais contém o intervalo interquartil dos valores de qualidade para cada base, mensuradas entre 25% e 75%, onde também pode-se ver linhas em preto acima e abaixo dessas colunas que correspondem aos valores 10% e 90% dos dados. Dispostas dentro de cada barra amarela há uma linha em azul, que consiste na flutuação do valor médio da qualidade e uma linha em vermelho, determinando o valor mediano de qualidade de cada base. **A)** Análise de qualidade realizada após a trimagem dos dados para a biblioteca L-laeta_S1_L001_R1_001.fastq. **B)** Análise de qualidade realizada após a trimagem dos dados para a biblioteca L-laeta_S1_L001_R1_001.fastq.

Na Tabela 3, pode se verificar que ao todo foram eliminadas 13.33 % de todas as *reads* sequenciadas do arquivo R1 e 14,05 % das *reads* do arquivo R2, dentre elas estão as *reads* com tamanho inferior a 40 pb e sequências que não possuem qualidade superior a Q20 em toda a extensão da *read*. Foi recuperado um número considerável de *reads* com qualidade superior a

Q20 em toda a extensão da *read* e com um tamanho médio de 286 pb Concomitantemente, 68,16% das bases do arquivo R1 foram preservadas e do mesmo modo 60,82% bases do arquivo R2.

Tabela 3 Relação quantitativa de sequências geradas antes e após trimagem.

Input	Arquivo Inicial		Arquivo final		% reads após a trimagem	% bases após a trimagem
	Reads	Bases	Reads	Bases		
L-laeta_S1_L001_R1_001.fastq	11 870 865	2 392 153 725	10 289 208	1 630 584 144	86,67	68,16
L-laeta_S1_L001_R2_001.fastq	11.870.865	2 469 510 306	10 203 328	1 502 190 333	85,95	60,82

Pelo o que foi observado durante o processo de tratamento das sequências, quando se utiliza os parâmetros padrões de trimagem do Prinseq, existe o risco de que inúmeras sequências sejam negligenciadas e descartadas do arquivo final. Isso forma um conjunto de dados mais escasso e diminui a possibilidade de uma montagem mais completa. Nesse trabalho, executamos um processo de trimagem criterioso, para que pudesse ser elevada a qualidade do arquivo como um todo sem que houvesse descarte de um número grande de *reads*. Durante o processo de trimagem, a variação na retirada das sequências a partir da extremidade 3' com valores de qualidade entre 22 ao 36, concomitante a variação da janela entre 50 e 100 nucleotídeos, permitiu a manipulação do arquivo de maneira que apenas a maioria das bases de baixa qualidade foi removida. Assim, foi mantido um grande número de informações sequenciais que funcionaram como subterfúgio para as saídas consenso durante a montagem posterior.

5.3 Montagem *de novo*

As *reads* de trimadas foram utilizadas para montagem, através do programa Trinity, de um único transcriptoma *de novo* de *L. laeta* peruana. Foram realizadas 10 montagens *de novo* para o mesmo transcriptoma de *L. laeta*, com variações de K-mer de 18 ao 32. A montagem mais substancial e expressiva do transcriptoma apresentou um total de 143.237 contigs.

Através de alinhamentos de cada uma dessas montagens contra o banco de dados para fosfolipases do NCBI, foi possível constatar que a montagem com valor de K-mer 22 apresentou cobertura melhor de seus contigs contra o conjunto de dados para Fosfolipases D. Foram encontradas mais correspondências entre o cruzamento das sequências na montagem com K-mer 22 e os valores de similaridade e identidade apresentaram-se mais altos em comparação às outras montagens. Isso revela que a escolha do tamanho k-mer tem resultados distintos na formação de consenso entre as sequências (COMPEAU, P., et al., 2011; DURAI AND SCHULZ, 2016). Tamanhos inferiores de k-mer diminuem a quantidade de arestas armazenadas no gráfico do algoritmo *de Bruijn*, e os superiores irão aumentar (ZERBINO e BIRNEY, 2008).

A obtenção de mais de uma montagem além da configurada segundo os padrões do *software* possibilita uma diversidade maior de dados e possibilidade de encontrar novas toxinas não cobertas pelos parâmetros tradicionais. O número de *contigs* obtidos foi amplo e possibilita abranger futuros estudos sobre o transcriptoma da *L. laeta* peruana.

A montagem de transcriptomas com RNA-Seq viabilizou a identificação de possíveis transcritos raros que não seriam encontrados em bibliotecas clássicas de cDNA. Isso se deve à otimização do processo e ao fato de que são gerados um número muito maior de dados do que o método antigo, o que favorece a descoberta de sequências nunca antes estudadas.

5.4 Visão geral das fosfolipases no transcriptoma da peçonha Loxoscélica

As fosfolipases são os principais componentes da peçonha das aranhas marrons. O presente estudo visou analisar as diferenças e similaridades em relação às fosfolipases D presentes em peçonha de aranhas do gênero *Loxosceles*. A partir da análise e descrição da expressão de toxinas da peçonha além de estudos de interatividade molecular, torna-se possível o aperfeiçoamento de ferramentas para o estudo e o aprofundamento dos conhecimentos dos mecanismos de ação dos diferentes constituintes das peçonhas, desde a descrição bioquímica estrutural e funcional, até os efeitos biológicos e farmacológicos.

A identificação e caracterização de proteínas da glândulas de peçonha é utilizada em muitos trabalhos com o propósito de estabelecer uma relação entre os peptídeos e a similaridade, ou seja, clusters de peptídeos que sejam mais semelhantes uns aos outros. Com isso, podemos então selecionar e avaliar os transcritos de uma peçonha, obtendo informações significativas sobre suas estruturas e funções. O arquivo em formato fasta, contendo as sequências montadas,

foi submetido ao alinhamento de sequências com a ferramenta BLASTx *stand-alone*, utilizando a base de dados de proteínas não-redundantes do GenBank.

O conjunto de dados, com os contigs do transcriptoma foi submetido a um alinhamento, através da Blastx *stand-alone*, contra o banco de dados de proteínas do NCBI relacionadas a fosfolipases. Uma quantidade significativa de contigs apresentaram similaridade, contudo os resultados que não se enquadravam nos parâmetros pré-determinados pelo estudo foram filtrados.

Dentre o resultado do alinhamento entre a montagem e o banco de dados do GenBank, dos 2593 *contigs* com hits significativos (*e-value* menor do que 10^{-4}), 1424 são de Fosfolipases-A, 315 de Fosfolipases-B, 274 de Fosfolipases-C, 453 de Fosfolipases-D e 127 são *contigs* com similaridade a sequências de fosfolipases não classificadas do banco de dados (Figura 11). A elevada proporção de hits alinhados com fosfolipases A, B e C pode ser explicada pela presença de fosfolipases de metabolismo, devido a técnica utilizada para a obtenção do RNA total. A maceração das glândulas pode levar transcritos do metabolismo epitelial ao conjunto da amostra, fazendo com que outras fosfolipases possam ser encontradas um número maior que as fosfolipases D.

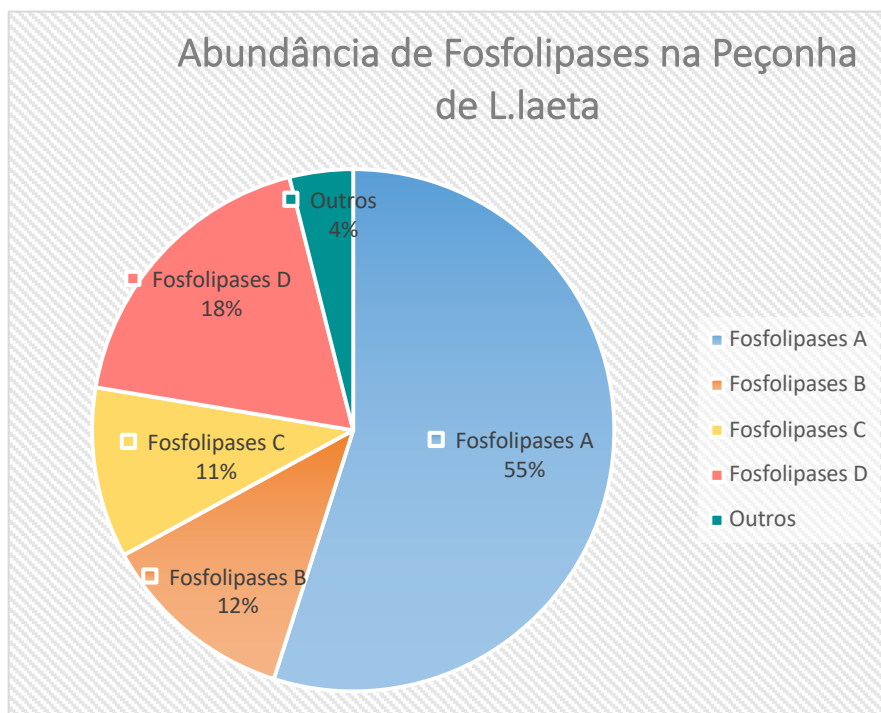


Figura 11 Abundância de fosfolipases no transcriptoma da glândula de peçonha de *Loxosceles laeta*. Fosfolipases D: 18% dos transcritos alinhados contra o banco de dados. Fosfolipases C: 11% dos transcritos. Fosfolipases A: 55% dos transcritos. Fosfolipases B: 12% dos transcritos. Outros: 4%.

Em seguida, foi realizada mais um processo de filtragem a fim de que permanecessem os alinhamentos com maiores resultados possíveis para posterior anotação com foco nas fosfolipases D (PLDs). Foram selecionados os *contigs* com maior cobertura e maior número de aminoácidos idênticos a proteína do banco de dados. Os alinhamentos que apresentaram proteínas que alinharam com *contigs* diferentes foram desempatados através da comparação entre as porcentagens de cobertura do alinhamento.

5.5 Anotação de Isoformas de Fosfolipases - D da peçonha de *L. laeta*

5.5.1 Identificação de sequências com similaridade às proteínas do banco de dados NCBI

Através da ferramenta Blastx *stand-alone*, segundo os parâmetros descritos no item 4.6, foram identificadas proteínas referente aos *contigs* do transcriptoma da glândula de peçonha da aranha-marrom peruana que apresentaram maior similaridade com foco nas PLDs. Foi observado um número elevado de proteínas do transcriptoma, com porcentagem de similaridade variada, contra as proteínas do banco dados GenBank. As proteínas do transcriptoma com melhores resultados do Blast foram mantidas e de acordo com os critérios predispostos no item

4.6 foram filtradas de modo que as sequências com maiores valores de similaridade e identidade fossem mantidas.

Para certificação dos resultados obtidos através do alinhamento do Blastx *stand-alone*, e após a análise e triagem dos resultados obtidos, as sequências dos contigs selecionadas (aproximadamente 100) foram levadas ao programa Blastx on line. Nesse passo, algumas proteínas não foram caracterizadas como PLDs completas ou alinharam com proteínas quiméricas, devido a isso, mantiveram-se as proteínas referentes aos contigs de *L. laeta* peruana completas e correspondentes a proteínas identificadas experimentalmente.

Nesse trabalho, foram descritas 19 sequências referentes aos *contigs* precursores das proteínas completas de isoformas de fosfolipases D potencialmente encontradas na peçonha de *L. Laeta* (Anexo 1). Foi possível através da utilização das informações obtidas sobre esses alinhamentos estimar a relação das sequências determinadas entre a classe e família dos peptídeos alinhados com essas sequências. Os transcritos encontrados foram submetidos a nova análise de similaridade através da plataforma online do programa Blastx contra o banco de dados geral on line do NCBI e todos os contigs submetidos a essa análise apresentaram similaridade a alguma fosfolipase descrita, principalmente as da superfamília PI-PLCc_GDPD_SF e da subfamília (GDPD_like_SMaseD_PLD).

Neste trabalho, encontramos uma relação significativa entre as PLDs correspondentes aos *contigs* advindos do transcriptoma de *L. laeta* peruana e as sequências de FLDs anteriormente relatadas de *L. laeta*. Além disso, houve similaridade dessas mesmas sequências com outras já relatadas do gênero *Loxosceles* (*L. intermedia*, *L. gaucho*, e *L. reclusa*, *L. arizonica*, *L. amazonica*, *L. rufescences*) e algumas espécies do gênero *Sicarius* (*S. damarensis*, *S. albospinus*).

5.5.2 Processamento dos *contigs* selecionados: tradução, identificação do peptídeo sinal, do ponto isoelétrico (P.I.) e massa molecular (m.m)

As fosfolipases D mais expressivas e com melhores valores selecionadas (Tabela 3) foram avaliadas quanto a presença de peptídeo sinal (através da ferramenta on line SignalP 4.1 Server), ponto isoelétrico (P.I.), Massa Molecular (M.M.), características da família e domínios conservados. Todas elas apresentaram em suas sequências protéicas as regiões de peptídeo sinal.

As sequências de cDNA dos *contigs* selecionados foram levadas ao programa que realiza tradução de moléculas de cDna para proteínas, Expasy, conforme descrito no item 4.6. As sequências completas resultantes da tradução podem ser encontradas no Anexo 2. A

ferramenta Expasy também ofereceu a possibilidade de delimitarmos a sequência relacionada a proteína madura, que foi utilizada nos alinhamentos a serem vistos adiante.

Para a identificação do peptídeo sinal, os contigs traduzidos foram submetidos a execução do programa SignalP 4.1. Observou-se que algumas proteínas apresentavam o mesmo peptídeo sinal, contudo sequências de aminoácidos diferentes.

Através da ferramenta do programa BlastP foi possível identificar que todas as proteínas descritas neste trabalho possuem um loop catalítico, um sítio ativo composto duas histidinas e um triptofano e um domínio de ligação ao magnésio (Mg^{2+}), que podem ser observados na figura 10.

Essas sequências foram identificadas como similares a fosfolipases D, esfingomielinases D ou proteínas dermonecróticas de aranhas do gênero *Loxosceles*, o que inclui: cinco *contigs* transcritos para a proteína de *L. laeta* (TAMBOURGI et al, 2008; PEDROSA et al., 2002); um transcrito para a *L. boneti* (RAMOS-CERRILLO E COLS., 2004); três transcritos para a de *L.reclusa* (LEE E LYNCH, 2005) e um transcrito para *L. amazonica*.

Também foi encontrada similaridade no alinhamento realizado entre proteínas de PLDs de outras espécies de gêneros diferentes, como *Sicarius*, contudo consideradas próximas evolutivamente (PLATNICK, 2016). Apesar de que as aranhas do gênero *Loxosceles* sejam mais conhecidas pela picada responsável por lesões dermonecrótica em tecidos de mamíferos, também há relatos de picadas de algumas espécies de *Sicarius* também causarem dermonecrose. Isso pode explicar a similaridade entre as PLDs do transcriptoma de *L.laeta* e algumas isoformas das PLDs do gênero *Sicarius* (NEULANDS E ATKINSON 1988 ; VAN ASWEGEN et al.1997).

O ponto isoelétrico e a massa molecular das fosfolipases identificadas foram avaliados com a ferramenta Compute pI/Mw da plataforma on line Expasy. Os valores encontrados para as proteínas maduras estão na Tabela 4. Todas as fosfolipases apresentam caráter ácido, com ponto isoelétrico em média de 7, com exceção dos DN5205_c0_g1_i1, DN625_c0_g1_i1, DN101041_c0_g1_i1, DN28139_c0_g1_i1 que possuem esse caráter levemente básico. As PLDs encontradas, apresentam massa molecular em média de 35 kDa, na forma madura.

Binford et al, 2008, estudaram também uma isoforma de PLD de uma espécie *L. laeta* encontrada Los Angeles, CA, EUA, que apresentou atividades dermonecróticas e sítio ativo correspondente ao sítio da isoforma do peptídeo DN25784_c0_g1_i1 (com 99.28 % de similaridade e 99.28 % de identidade) e com DN405_c0_g1_i1 (com 99.28 % de similaridade e 99.28 % de identidade), ambas de *L. laeta* peruana. Catalán et al, 2011, observaram que a proteína recombinante rLIPLD1 referente a Fosfolipase D isoforma 1 de *L. laeta* tinha atividade

hidrolítica em esfingomiéline e atividade hemolítica *in vitro* em eritrócitos do sangue humano. A PLD desse estudo apresentou alinhamento com 88.75% de identidade e 95.82% de similaridade com o peptídeo DN51055_c0_g1_i1. Fernandes Pedrosa e colaboradores, em um estudo de 2002, realizou a partir de uma biblioteca de cDNA com transcritos da glândula da peçonha de *L. laeta* brasileira a identificação, clonagem e expressão de PLDs mais encontradas nesse fluido. Os autores encontraram proteínas com sequências semelhantes a PLD da *L. laeta* brasileira (gi|27372516|) com a de *L. intermedia*, tamanho médio de 33kDa e características funcionais similares. As propriedades biológicas atribuídas a peçonha foi as pertencentes a família de esfingomielinases de *L. laeta* e de *L. intermedia*, que inclui atividades hemolítica, dermonecrótica e dependentes de sistema de complemento. Neste estudo, a proteína DN337_c1_g1_i1 apresentou melhor alinhamento com essa isoforma relatada por Fernandes Pedrosa, et al, a qual teve características físico químicas similares com a proteína descrita anteriormente.

Ramos-cerrillo et al, 2004, fez um estudo com espécies de *L. boneti* coletadas na cidade de Tuxpan, região de Guerrero, no México. Nesse estudo, os pesquisadores identificaram que as proteínas correspondentes a *L. boneti* de PLD apresentaram pesos moleculares de aproximadamente 30 kDa por SDS-PAGE. O que se assemelha ao peso molecular da proteína DN100733_c0_g1_i1 de PLD de *L. laeta* peruana identificada neste trabalho. Os pesquisadores também identificaram os três principais picos do cromatograma *L. boneti* que se referiam como as isoformas 1, 2 e 3 respectivamente, de PLD da espécie. O autor porém revela que curiosamente essa isoforma não exibe a atividade dermonecrótica quando inoculado em coelhos em três vezes a dose das outras isoformas, embora a sua identidade para as outras isoformas da mesma espécie é suficientemente elevada para justificar a sua identificação como uma isoforma de PLD. Neste trabalho a proteína DN100733_c0_g1_i1 melhor se alinhou a isoforma 3, apresentando alto índice de similaridade, de identidade e cobertura. Podemos observar no Anexo 3 que as histidinas e cisteínas estão bem conservadas no alinhamento. Apesar do sítio ativo estar bem alinhado e ser caracterizado como região de atividade da enzima em outras isoformas e em PLDs de outras espécies, nesse caso seria interessante um ensaio experimental para verificação da atividade da isoforma 3 caracterizada de PLD de *L. laeta* peruana, bem como a possibilidade de existência de outros sítios ativos e se a enzima pode ter assumido outras funções perante as diferenças ambientais.

Lee e Lynch, 2005, realizaram um estudo que identificou uma proteína de PLD a partir das glândulas de peçonha de *L. reclusa* com 305 aminoácidos e massa molecular de 34 kDa, dados semelhantes aos peptídeo DN545_c0_g1_i1 e DN51591_c0_g1_i1 encontrados no

alinhamento entre as sequências de *L.laeta* peruana e o peptídeo relatado anteriormente. Segundo o autor, a sequência do peptídeo sphingomyelinase D precursor [*L. reclusa*] revelou um alinhamento de aminoácidos de 84% e 46% de *L. boneti* SMaseD1 e SMaseD2, respectivamente, que segundo Ramos-cerrillo et al, 2004 constituem figuras de caráter dermonecrótico na peçonha. A análise da sequência do peptídeo de *L.boneti* sugeriu um domínio de potencial catalítico HKD , responsável pela dermonecrose que foi diagnosticada experimentalmente com o efeito dermonecrótico da enzima, além de sinais de hemólise de eritrócitos.

Um estudo realizado por Binford et al, 2008, classificou a família de PLDs de *Sicarius* como sictox e observou que as proteínas expressas das *Sicarius* variavam entre 31 e 35 kDa. Os pesquisadores desse estudo, tiveram como objetivo analisar a evolução molecular e variação na atividade da família de PLDs entre os peçonhas brutas das espécies *Loxosceles* e *Sicarius* de maior incidência na América e África. Dentre as diversas espécies por ele estudadas, foi identificada a proteína sphingomyelinase D-like protein, partial, relacionada a *Sicarius albospinus* nativa da região Namibia: Gobabeb, na Africa e a proteína sphingomyelinase D-like protein, partial relacionada a *Sicarius damarensis*. Essas sequencias apresentaram homologia significativa com espécies de *Loxosceles* de outras regiões, inclusive América do Sul e nesse caso apresenta atividade dermonecrótica diferente de muitas espécies de *Sicarius* da América. Essas proteínas foram relatadas como sequências parciais de fosfolipase D no estudo de Binford et al, 2008, e curiosamente apresentaram similaridade de 84.81% e 81.39% no alinhamento com as sequências DN125979_c0_g1_i1 e DN5205_c0_g1_i1 de *L.laeta* peruanas, respectivamente. Apesar dos hits serem de sequências parciais, a proteína de *L.laeta* peruana encontrada no transcriptoma se caracteriza por ser total, baseada na presença e definição do peptídeo sinal e identificação da proteína madura.

Através da ferramenta para análise bioinformática *Decrease Redundancy* (http://web.expasy.org/decrease_redundancy/), foram mensuradas as porcentagens de redundâncias das sequências protéicas traduzidas a partir das sequências de cDNA dos *contigs*, para exclusão de possíveis duplicatas. Observou-se que os *contigs* DN101856_c0_g1_i1 e DN101111_c0_g1_i1 apresentaram 100% de identidade entre si. Contudo, é interessante mencionar que ambas possuem cDNA diferentes (Anexo 1), contudo codificam a mesma proteína e por isso os dois foram citados na tabela 4. Com fins didáticos, para a realização dos alinhamentos entre as sequências protéicas dos *contigs* e as proteínas selecionadas do banco de dados, foi mantida apenas uma dentre essas duas proteínas.

DN100812_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 74814733 sp Q8I913.1 [Loxosceles laeta]	MYVHLAIIIGCWTVVLOGAETD VGGERADNRRIWNL GHMVNAVVKQIPTFLNDGANAIEADITFKGAVPTYSYHG TPCDFGRDCIRWEYFDVFLQTLRDYTPPGNSKYHEKFIL FVLDLKTGSLNNEVRKAGENVAKGLLQNYWNNGNN GGRAYVVLSPDIAHYEFIRTFKEVLKAEGHENLLDKV GYDLSPYLPSPSLDSVHEAFKAGVDGHVWLSDDL TNWAPLGD MARLKEIVKRRDSENGFISKVYVYWSVDKY STTRTALDVGVDGIMTNFPYVIDVLNEDGYKDNRYLA TYDDNPWETFKK-	0.0	5.52	35285.79
DN51591_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 209573787 LOXIN [Loxosceles intermedia]	MIFNIVLTIGFLAVLLEGAETD VSERADNRRIWNNMG HMVNAIQIDEFVDLGANSLEIDITFDKKAKADYTYHKL PCDCFRTRCKWEYVNDYLVKAVRRATTPGYSKYRSRLV LLVFDLKTTLTNTAYKAGEDFARRLLEHYWNNGND GGRAYIVLSIPDVNHYTFINGFGEALRSSGHPQLMQKV GYDFSGNDDLSTIRNALRKAGVNDKEHIWQSDGITNCL ARGLSRVREAVNNRDSNGYISKVYVYWTVDKYKSIRD ALDAGVDGVMNTNPNRVIDVLNENGYSNRFRLATYSD NPFETF-	4E-160	7.01	35025.37
>DN25784_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 292495526 LOXLA[Loxosceles laeta]	MLLDISLLLGFCVAVLLOGAETE VIERADNRRIWNNMG HMVNEIQIDEFVDLGANSIETDITFDDDAIAEYTYHGV PCDCRWCWKWENVNDYLVHGLQRATTPGNSKYHPELV LVVFDLKTGDLSSSTAYKAGNMFAQKLFHYWNAGNN GGRAYIVLSIPDIDHYAFISGFREAFKNSDHADLLDKVG YDFSGNDDL SATRNALNKGKGVKDRHVWQSDGITNCL GRGLGRVRDAVANRDSNGYINKVYVWVWIEKYVSRVD AFDAGVDGIMTNEPDVIVDLNESAYSSKFRMATYED NPWETFKYK-	0.0	4.86	34851.91
DN100733_c0_g2_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 118573878 [Loxoscel es intermedia]	MWLSAVISFIGFAVFLQEA NGYVEERADSRKPIWDIA HMVNDLALVDEYLDGAGANALEADLAFKSDGTAEKMY HGTPCDFRSCSRTEKFDKYMDYIREITTPGSSKFKSGM LLIIDLKLNIEPNVAYAAKSTAEITLLKNYWQNGES KARAYIVLSLETITRPFDFISGFKDAIDASGHTELYEKIGW DFSGNEDLDIRRAYQKYGIDDHIWQGDGITNCLPRGD YRLTEAMKKNNDPKYTKKVVYTSIDKESIRNALR LGVDAMVMTNYPARVKSVLQESEFGSHRMATYDDNP WVK-	0.0	5.29	34431.79
DN101786_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 121987412 sp Q27Q5 4.1 [Loxosceles intermedia]	MLLSILYCLIRPSILLDA ANGEGFMERAVDGRRIWNL IAHMTNDIEIVRTYLNAGANALEFDIEFNKNGKPVRTFH GSPDCCRKCDRTTTFSTFIEQMRKFTTPGDPEFKKQLV LLMLDLKVTDFDKSLAYTAGQNVAEVLLDNFWQRGK SKPQAYLLSIPSLDYAKLIKGFNDKLDKDEGFVEVYQEK VGVDFSNNDLDDVRKTLKSIGFESRIWQSDGITNCLPR LTGRLKEAIQRDTPGYTYINKVYVWTLDRKSTMRSTL DLGVDMITNYPDLNLSVLEEEYVSRFRLATVEDNP WEIFNPSGSRMIEESYDSSLDEEYDCLKKRETQDFIPFF LGELFASCAGKALF-	2e-88	4.96	41099.72
DN82077_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 292630589 sp C0JB8 0.1 [Sicarius cf. Damarensis]	MLLSILYCLIRPSILLDA ANGEGFMERAVDGRRIWNL IAHMTNDIEIVRTYLNAGANALEFDIEFNKNGKPVRTFH GSPDCCRKCDRTTTFSTFIEQMRKFTTPGDPEFKKQLV LLMLDLKVTDFDKSLAYTAGQNVAEVLLDNFWQRGK SKPQAYLLSIPSLDYAKLIKGFNDKLDKDEGFVEVYQEK VGVDFSNNDLDDVRKTLKSIGFESRIWQSDGITNCLPR LTGRLKEAIQRDTPGYTYINKVYVWTLDRKSTMRSTL DLGVDMITNYPDLNLSVLEEEYVSRFRLATVEDNP WEIFNPSGSRMIEESYDSSLDEEYDCLKKRETQDFIPFF LGELFASCAGKALF-	3e-95	4.96	41099.72
DN28139_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 292630585 sp C0JB7 2.1 [Sicarius cf. damarensis]	MFSATAYLVALSIVLOGTQKY FEADSKRPFWNIAHM VNAIDQINPFLKRGANAIEFDIVFNNGIAEQTHHGVP DCGRICRTRKEDFVKYLDHIRLVTPDSKEFRDQLLLAL DLKLQRISTSKAYAAGEDIANKLIDHYWQKNGTARA YILLNIPSIRHDFINGFQHTIIRREGYERYNDKFGINFTG NDDLEATRMLKRMNITSNIWQADGITSFPRGTRRLK DAIRRRDTPGYKFIKVVYSWTLVTFAMRRSMRLGVDG IMTNHPERVAYILGEKYFANRFRLATIEDNPWQKYQP-	3E-117	9.59	35545.81
DN125979_c0_g1_i1 i1 (PI- PLCc_GDPD_SF Superfamily)	gi 292630598 sp C0JB9 1.1 [Sicarius albospinus]	MLFCTILCIAGLSVIEA AESISAERDDAKRPFYNIAH MVNGIVQINQFLDKGANAEFDIEFNNDGIAQITHHGFP CDCGRKCSFQENVTEYLDHIRQVTPGDPKFREQLVLL ALDLKLQRIFSKAHTAGVDVAKKLLDHYWQRGKSK ARAYLLNIPLLKDYEFVRGFKETLRDEGYEQYNDVIGI NFTGNEDLDKTRREIYQKLNVSIIHWQSDGITSFRRKTK RLKEALNLRDDPNYNYVTKVYVWTLVKYTSMKERLL DVGDIQTNYPDLRDEVLKEKEFASMFRLATFEDDPWK KFMRA-	2E-131	6.67	35892.97

DN625_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 122108190 sp Q1W6 94.1[Loxosceles intermedia]	MOLFIALCLLSSIGLLEG AKQDGIERTDSGRPIWNIAH MVNDKGLIDEYLAAGANSVESDVSFDDNGKPMKFHH GRPCDCGRNCDRQMSFTDYLDYMRQLTTPGTDNTMFR ENLILIMLDFKLLKLSPNVAYTAGQEVANQMLDITYWK RGKSEARAYFVLSIPSIKKAFAKGFYDKLNSEGFDQYR EKVGVDFSGNEAFADTRATLSSLNIVDHIWQSDGITNC LPRNMKRLKKAVKERDSNGYMKVYNWSVDKYTTM RKVLRAGADGMITNFPGRLLVSVLNREFSGKFRMATY DDNPWEKY-	2E-172	8.44	34736.52
DN101041_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 122108190 sp Q1W6 94.1[Loxosceles intermedia]	MOLFIALCLLSSIGLLEG AKQDGIERTDSGRPIWNIAH MVNDKGLIDEYLAAGANSVESDVSFDDNGKPMKFHH GRPCDCGRNCDRQMSFTDYLDYMRQLTTPGTDNTMFR ENLILIMLDFKLLKLSPNVAYTAGQEVANQMLDITYWK RGKSEARAYFVLSIPSIKKAFAKGFYDKLNSEGFDQYR EKVGVDFSGNEAFADTRATLSSLNIVDHIWQSDGITNC LPRNMKRLKKAVKERDSNGYMKVYNWSVDKYTTM RKVLRAGADGMITNFPGRLLVSVLNREFSGKFRMATY DDNPWEKY-	9e-174	8.65	34702.49
DN101111_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 57792507 gb AAW56 831.1 (Loxosceles reclusa)	MFLYVLLGVSCWILILKGTEA DSDEVGDRRRPIWNIA RTVNAEYPIDMFLEGDVSPIGITEFLDLGANAIQLEVIFD SSGNAESVYQKTPCDCRRTCVWRWRDFSYYLDDLRRFT TPGELTFREELIHLIDLKVDGFNSLSYTAGKNLAEKLL ENYWQEENTESNARAYILLSVPSTKNAALFRAFNLTLQ NGGFAAYGEKVGINFAEDDDLFVSVRIVNRLGTAGLHV WQSSVITNCLLRENSELSDAVQERDSQGYMKNVYTLT ADRHATIRNALRASVDGVMNTNPNHMADVLKENEFKD KFKLATYDDNPWKTYVEETEDVSNELDVDDVNNEEVI-	1,00E-84	4.52	38709.34
DN101856_c0_g1_i1 (PI- PLCc_GDPD_SF Superfamily)	gi 57792507 gb AAW56 831.1 (Loxosceles reclusa)	MFLYVLLGVSCWILILKGTEA DSDEVGDRRRPIWNIA RTVNAEYPIDMFLEGDVSPIGITEFLDLGANAIQLEVIFD SSGNAESVYQKTPCDCRRTCVWRWRDFSYYLDDLRRFT TPGELTFREELIHLIDLKVDGFNSLSYTAGKNLAEKLL ENYWQEENTESNARAYILLSVPSTKNAALFRAFNLTLQ NGGFAAYGEKVGINFAEDDDLFVSVRIVNRLGTAGLHV WQSSVITNCLLRENSELSDAVQERDSQGYMKNVYTLT ADRHATIRNALRASVDGVMNTNPNHMADVLKENEFKD KFKLATYDDNPWKTYVEETEDVSNELDVDDVNNEEVI-	3,00E-85	4.52	38709.34

5.5.3 Alinhamento entre as sequências de proteínas completas do transcriptoma da Aranha-marrom peruana e proteínas previamente descritas

Para a realização do alinhamento múltiplo entre *contigs*, além de alinhados pelo Blastx *stand-alone*, foram alinhados também na plataforma on line do Blastx para verificação de maior variedade de sequências de *L.laeta*, proporcionando enriquecimento do alinhamento e possibilidade de comparação com mais isoformas anteriormente estudadas. Com as sequências de proteínas similares inferidas, foi então utilizado o programa ClustalW para obtenção do alinhamento.

O alinhamento de maneira geral entre as proteínas dos *contigs* (figura 12), mostrou a similaridade entre as sequências dos *contigs*. Também é possível observar resíduos de cisteínas conservados, que possivelmente formam duas pontes dissulfeto, além do sítio ativo formado pelo domínio H também está presente nas sequências e o resíduo de metionina localizado na alça hidrofóbica. O sítio catalítico, o domínio de ligação ao Mg²⁺ e o loop catalítico foram as regiões que apresentaram maior índice de conservação. Todos as isoformas, com exceção das isoformas DN101856_C0_g1_i1 e DN51608_C0_g1_i1, apresentaram sítio catalítico e

domínio de ligação ao Mg²⁺ conservados. A proteína DN101856_C0_g1_i1 possui uma arginina e uma glicina substituindo histidinas consideradas importantes na atividade das PLDs. Também identificamos duas mutações no domínio de ligação ao Mg²⁺, nas quais um glutamato foi substituído por uma glutamina e um aspartato foi substituído por um glutamato. A proteína DN51608_C0_g1_i1 possui uma asparagina e um glutamato substituindo as histidinas consideradas importantes na atividade das PLDs. Apesar disso, o domínio de ligação ao Mg²⁺ foi conservado.

Foi realizado também um alinhamento entre as proteínas do transcriptoma da aranha *L. laeta* peruana e as isoformas de aranhas *L. laeta* encontradas no Brasil para comparação das PLDs entre as espécies. Nos resultados obtidos que estão representados na figura 13, verifica-se a existência de homologia entre todas as proteínas, com zonas muito semelhantes. Constatamos características importantes que confirmam a relação dessas sequências no transcriptoma entre as PLDs já descritas na literatura, como por exemplo o grau significativo de conservação das cisteínas. Do mesmo modo, pôde ser inferido a conservação dos sítios ativos das proteínas correspondentes aos *contigs*, as histidinas.

Foram determinadas para participar do alinhamento contra as PLDs do transcriptoma, proteínas representantes de cada grupo funcional delimitados no cladograma da figura 14, representando um terceiro alinhamento. Podemos comparar agora algo peculiar entre as sequências que pode ser objeto de futuros estudos, devido a presença de regiões de mutação nos sítios ativos de histidina das isoformas DN101856_C0_g1_i1 e DN51608_C0_g1_i1. Através do alinhamento entre as sequências do transcriptoma contra sequências do banco de dados NCBI, observamos que tal mutação, referente a isoforma DN101856_C0_g1_i1 também já foi identificada em outra proteína da espécie *L. similis*, anteriormente relatada (DANTAS et al., 2016), o que pode ser visto no Anexo 3. Essas proteínas, podem apresentar uma baixa ou ausência de atividade enzimática em mamíferos, pelo fato de apresentar mutações em aminoácidos importantes para sua função (VUITIKA et al., 2013). Com isso, a DN101856_C0_g1_i1 torna-se um alvo de estudos para seu uso na imunização para a produção de soro, visto que podem funcionar como epítomos eficientes para a geração de anticorpos neutralizantes (CHAVEZ-OLÓRTEGUI et al., 1997; MARTIN-EAUCLAIRE et al., 2006; MENDES ET AL., 2013; MOREIRA-FERREIRA et al., 1998).

10 20 30 40 50 60

DN101856_c0_g1_i1/1-341 -MFLYVLLGVSCWILILKGT EADS--DEVGD RRRPIWNIA RTVNAEYPI DMFLEGDVSP IGI TEF LD 64
 DN100733_c0_g2_i1/1-304 -MWLSAVISF IGF AVFLQEANGYV--EERADSRKPIWDIAH MVNDLALV-----DEYLG 51
 DN337_c1_g1_i1/1-317 -MYVHLALILGCWTVILQGA VTDV--GERADNRRPIWNLAH MVNAVVKQI-----P TFLD 51
 DN100812_c0_g1_i1/1-311 -MYVHLALILGCWTVV LQGAETDV--GERADNRRPIWNLGH MVNAVVKQI-----P TFLN 51
 DN51055_c0_g1_i1/1-311 -MYVHLALILGCWTVILQGA VTDV--GERADNRRPIWNLGH MVNAVVKQI-----P TFLN 51
 DN51608_c0_g1_i1/1-324 -MVF LVALMLG SWAVVLESGTEA--IKSADERQPIWNIANMV TDINQI-----NETLD 51
 DN100773_c0_g1_i1/1-305 -MLPYIALVLGCVGWVLES EAQTDV--GERADKRRPVWIMGH MVNDPPQI-----DEFVN 51
 DN25784_c0_g1_i1/1-309 -MLLDISLLLGFCAVLLQGAETEV--IERADNRRPIWNMGMH MVNEIYQI-----DEFVD 51
 DN405_c0_g1_i1/1-309 -MLLDISLLLGFCAVLLQGAETEV--IERADNRRPIWNMGMH MVNEIYQI-----DEFVD 51
 DN545_c0_g1_i1/1-306 -MIFNIVLTIGFLAVLLEGAETDV--SERADNRRPIWNMGMH MVNAIYQI-----DEFVD 51
 DN51591_c0_g1_i1/1-306 -MIFNIVLTIGFLAVLLEGAETDV--SERADNRRPIWNMGMH MVNAIYQI-----DEFVD 51
 DN125979_c0_g1_i1/1-308 -MLFCTILC IAGLSVVI EAAESIS--AERDDAKRPFYNI AHMVNGIVQI-----NQFLD 51
 DN5205_c0_g1_i1/1-305 --MFSATAYLVALSIVLQGTQ--K--YVFADSKRPFWNI AHMVNAIDQI-----N PFLK 48
 DN28139_c0_g1_i1/1-305 --MFSATAYLVALSIVLQGTQ--K--YVFADSKRPFWNI AHMVNAIDQI-----N PFLK 48
 DN82077_c0_g1_i1/1-358 MLLSLLVLC LIRPSILLDAANGEGFMERAVDGR RPIWNI AHMTNDIEIV-----RTYLN 54
 DN101786_c0_g1_i1/1-358 MLLSLLVLC LIRPSILLDAANGEGFMERAVDGR RPIWNI AHMTNDIEIV-----RTYLN 54
 DN101041_c0_g1_i1/1-305 -MQLFIALCLLLS IGLLEGAQDQIE--RTDSGRPIWNI AHMVNDKGLI-----DEYLA 51
 DN625_c0_g1_i1/1-305 -MQLFIALCLLLS IGLLEGAQDQIE--RTDSGRPIWNI AHMVNDKGLI-----DEYLA 51

70 80 90 100 110 120 130

DN101856_c0_g1_i1/1-341 LGANAL EQL EVIDSSGNAESVYQKTPCDDRRITCVWRDFSYLLD LRRFTTTPGEL--TFREEL L I I 129
 DN100733_c0_g2_i1/1-304 DGANAL EADLAFKSDGTA EKMYHGTPCDDFRSCSRTEKFDK YMDYIREITTPGSS--KFKSGML LLI 116
 DN337_c1_g1_i1/1-317 LGANAL EADVTFK-GSVPTTYHGTPCDDGRDCIRWEYFNVFLQTLRDYITTPGNA--NYREQF L LFV 115
 DN100812_c0_g1_i1/1-311 DGANAL EADITFK-GAVPTYSYHGTPCDDGRDCIRWEYFDFVFLQTLRDYITTPGNS--KYHEKFI L LFV 115
 DN51055_c0_g1_i1/1-311 DGANAL EADITFK-GAVPTYSYHGTPCDDGRDCIRWEYFNVFLKTLREYITTPGNA--KYRDGF L LFV 115
 DN51608_c0_g1_i1/1-324 LGANSI ETDVTFDDGTMAGYTYEGVLCAGSGSC TKKMNIGEF LHGLRVRITTPGDP--MYRKE LALV 116
 DN100773_c0_g1_i1/1-305 LGVNSI ETDVSDNNANPEYTYHGTPCDDGRSCGHYEFDFVFLK NLRIRITTPGSS--KYHEKFI L LFV 116
 DN25784_c0_g1_i1/1-309 LGANSI ETDITD DDDAIAEYTYHGTPCDDGRWC TKWENVNDFLHGLQRAITTPGNS--KYHPELV L LV 116
 DN405_c0_g1_i1/1-309 LGANSI ETDITD DDDAIAEYTYHGTPCDDGRWC TKWENVNDFLHGLQRAITTPGNS--KYHPELV L LV 116
 DN545_c0_g1_i1/1-306 LGANSLE IDITD KKAADYTYHGTPCDDGRFTRCKWEYVNDY LKAVRRATTPGYS--KYRSRLV L LV 116
 DN51591_c0_g1_i1/1-306 LGANSLE IDITD KKAADYTYHGTPCDDGRFTRCKWEYVNDY LKAVRRATTPGYS--KYRSRLV L LV 116
 DN125979_c0_g1_i1/1-308 KGANAL EFDVFNDDGIAQI THHGFPCDDGRKCS FQENVTEYLDH I RQVITTPGDP--KFRQLV L LA 116
 DN5205_c0_g1_i1/1-305 RGANAL EFDVFN YNGIAEQTHHGFPCDDGRIC TRKEDFVK YLDH I RLVITTPDSK--EFRDQL L LA 113
 DN28139_c0_g1_i1/1-305 RGANAL EFDVFN YNGIAEQTHHGFPCDDGRIC TRKEDFVK YLDH I RLVITTPDSK--EFRDQL L LA 113
 DN82077_c0_g1_i1/1-358 DGANAL EFDIEFNKNGK PVRTFHGSPCDDGRKCDRTTTFSTFIEQMRKFTTPGDP--EFKQLV L LM 119
 DN101786_c0_g1_i1/1-358 DGANAL EFDIEFNKNGK PVRTFHGSPCDDGRKCDRTTTFSTFIEQMRKFTTPGDP--EFKQLV L LM 119
 DN101041_c0_g1_i1/1-305 AGANSV ESDVSD DNGKPMKFHHRPCDDGRNCDRQMSFTDYLDYMRQLTTPGTDN TMFREN L I IM 118
 DN625_c0_g1_i1/1-305 AGANSV ESDVSD DNGKPMKFHHRPCDDGRNCDRQMSFTDYLDYMRQLTTPGTDN TMFREN L I IM 118

140 150 160 170 180 190

DN101856_c0_g1_i1/1-341 LD LKVDGFNSLSYTAGKNLAEK LLENYQEEENT-----ESNARA Y I LLSV P STKNA A 182
 DN100733_c0_g2_i1/1-304 ID LK L N G I E P N V A Y A A C K S T A E T L L K N Y W Q N G -----E S K A R A Y I V L S L E T I T R P D 166
 DN337_c1_g1_i1/1-317 LD LK T G S L N N D E V R K A G E N V A K G L L E H Y W N N G -----N N G G R A Y V V L S L P D I A H Y E 166
 DN100812_c0_g1_i1/1-311 LD LK T G S L N N N E V R K A G E N V A K G L L Q N Y W N N G -----N N G G R A Y V V L S L P D I A H Y E 166
 DN51055_c0_g1_i1/1-311 LD LK T G S L S N D Q V R P A G E N V A K E L L Q N Y W N N G -----N D G G R A Y V V L S L P D I A H Y E 166
 DN51608_c0_g1_i1/1-324 F N L K T S D F S P S T A Y E S G M M F C G K L L Q I Y W N G G K T E R K C D P I K G I V V S R N D G G K A Y I I S I P D I E Y F A 183
 DN100773_c0_g1_i1/1-305 F D L K T G S L Y D N Q A N D A G K I A K S L L Q N Y W N N G -----N N G G R A Y I V L S I P H L K H Y Q 167
 DN25784_c0_g1_i1/1-309 F D L K T G D L S S S T A Y K A G N M F A Q K L F I H Y W N A G -----N N G G R A Y I V L S I P D I D H Y A 167
 DN405_c0_g1_i1/1-309 F D L K T T K L T T N T A Y K A G E D F A R R L L E H Y W N N G -----N D G G R A Y I V L S I P D V N H Y T 167
 DN545_c0_g1_i1/1-306 F D L K T T K L T T N T A Y K A G E D F A R R L L E H Y W N N G -----N D G G R A Y I V L S I P D V N H Y T 167
 DN51591_c0_g1_i1/1-306 LD L K L Q R I F S K A H T A G V D Y A K K L D H Y W Q R G -----K S K A R A Y V L L N I P L L K D Y E 167
 DN125979_c0_g1_i1/1-308 LD L K L Q R I S T S K A Y A G E D I A N K L I D H Y W Q K G -----N G T A R A Y I L L N I P S I R H F D 164
 DN5205_c0_g1_i1/1-305 LD L K L Q R I S T S K A Y A G E D I A N K L I D H Y W Q K G -----N G T A R A Y I L L N I P S I R H F D 164
 DN28139_c0_g1_i1/1-305 LD L K V T D F D K S L A Y T A G Q N V A E V L D N F W Q R G -----K S K P Q A Y L L S I P S L D Y A K 170
 DN82077_c0_g1_i1/1-358 LD L K V T D F D K S L A Y T A G Q N V A E V L D N F W Q R G -----K S K P Q A Y L L S I P S L D Y A K 170
 DN101786_c0_g1_i1/1-358 LD F K L K K L S P N V A Y T A G Q E V A N Q M L D T Y W K R G -----K S E A R A Y F V L S I P S I K K A S 169
 DN101041_c0_g1_i1/1-305 LD F K L K K L S P N V A Y T A G Q E V A N Q M L D T Y W K R G -----K S E A R A Y F V L S I P S I K K A S 169
 DN625_c0_g1_i1/1-305 LD F K L K K L S P N V A Y T A G Q E V A N Q M L D T Y W K R G -----K S E A R A Y F V L S I P S I K K A S 169

210 220 230 240 250 260

DN101856_c0_g1_i1/1-341 L F R A F N L T - L Q N G G F A A Y G E K V G I N F A E D D D - - - - - L F S V S R I V N R L G T A - G L H V W Q S S V I T N C L L R 242
 DN100733_c0_g2_i1/1-304 F I S G F K D A - I D A S G H T E L Y E K I G W D F S G N E D - - - - - L G D I R R A Y Q K Y G I D - - - - - D H I W Q G D G I T N C L P R 226
 DN337_c1_g1_i1/1-317 F I R A F K E V - L K R E G H E N L L E K V G Y D F S G P Y V P S L P T L D Q T H E A F K A G V E - - - - - G H V W Q S D G L T N F A P L 230
 DN100812_c0_g1_i1/1-311 F I R T F K E V - L K A E G H E N L L D K V G Y D L S G P Y L P S L P S L D S V H E A F K A G V D - - - - - G H V W L S D G L T N W A P L 230
 DN51055_c0_g1_i1/1-311 F I S G F R K T - L E I T G N E E I L A R V G Y E V S G N Y - - - - - L G L I H D A L R K A G V K D R E H V W Q S D G I T N C I P R 244
 DN51608_c0_g1_i1/1-324 L I T G F Q E T - L K T E G H E E L L D K V G Y D F S G N D - - - - - I S D V E N A Y K A G V T - - - - - G H V W Q S D G I T N C L T R 226
 DN100773_c0_g1_i1/1-305 F I S G F R E A - F K N S D H A D L L D K V G Y D F S G N D - - - - - L S A T R N A L N K G G V K D R E H V W Q S D G I T N C I G R 228
 DN25784_c0_g1_i1/1-309 F I S G F R E A - F K N S D H A D L L D K V G Y D F S G N D - - - - - L S A T R N A L N K G G V K D R E H V W Q S D G I T N C I G R 228
 DN405_c0_g1_i1/1-306 F I N G F G E A - L R S S G H P E L M Q K V G Y D F S G N D - - - - - L S T I R N A L R K A G V N D K E H I W Q S D G I T N C L A R 228
 DN51591_c0_g1_i1/1-306 F I N G F G E A - L R S S G H P E L M Q K V G Y D F S G N D - - - - - L S T I R N A L R K A G V N D K E H I W Q S D G I T N C L A R 228
 DN125979_c0_g1_i1/1-308 F V R G F K E T - L R D E G Y E Q Y N D V G I N F T G N E D - - - - - L D K T R E I Y Q L N V S - - - - - S H I W Q S D G I T S C F R R 226
 DN5205_c0_g1_i1/1-305 F I N G F Q H T I I R R E G Y E R Y N D K F G I N F T G N D - - - - - L E A T R R M L K R M N I T - - - - - S N I W Q A D G I T S C F P R 224
 DN28139_c0_g1_i1/1-305 F I N G F Q H T I I R R E G Y E R Y N D K F G I N F T G N D - - - - - L E A T R R M L K R M N I T - - - - - S N I W Q A D G I T S C F P R 224
 DN82077_c0_g1_i1/1-358 L I K G F N D K - L K D E G F E V Y Q E K V G D F S N N D - - - - - L D D V R K T L K S I G F E - - - - - S R I W Q S D G I T N C L P R 229
 DN101786_c0_g1_i1/1-358 L I K G F N D K - L K D E G F E V Y Q E K V G D F S N N D - - - - - L D D V R K T L K S I G F E - - - - - S R I W Q S D G I T N C L P R 229
 DN101041_c0_g1_i1/1-305 F A K G F Y D K - L N S E G F D Q Y R E K V G D F S G N E A - - - - - F A D T R A T L S S L N I V - - - - - D H I W Q S D G I T N C L P R 228
 DN625_c0_g1_i1/1-305 F A K G F Y D K - L N S E G F D Q Y R E K V G D F S G N E A - - - - - F A D T R A T L S S L N I V - - - - - D H I W Q S D G I T N C L P R 228

270 280 290 300 310 320 330

DN101856_c0_g1_i1/1-341 -E N S E L S D A V Q K R D S Q - - - - - G Y M N K V Y T L T A D R H A T I R N A L R A S V D G V M T N F N H M A D V L K E N E F K D K F 306
 DN100733_c0_g2_i1/1-304 G D Y R L T E A M K K R D P D - - - - - Y K Y T K V Y T W S I D K E S S I R N A L R L G V D A V M T N Y P A R V K S V L Q E S E F Y G K D Y 292
 DN337_c1_g1_i1/1-317 L N I D R L K E A I K S R D S A - - - - - N G F I N K I Y W S V D K Y S T T K A A L D V G V D G I M T N Y P Y V I I D V L N E D G Y K D Y 296
 DN100812_c0_g1_i1/1-311 G D M A R L K E I V K R R D S E - - - - - N G F I S K V Y W S V D K Y S T T R T A L D V G V D G I M T N F P Y V I I D V L N E D G Y K D N Y 296
 DN51055_c0_g1_i1/1-311 G D M A R L K E I V K R R D S E - - - - - N G F I S K V Y W S V D K Y S T T R T A L D V G V D G I M T N F P Y V I I D V L N E D G Y K D N Y 296

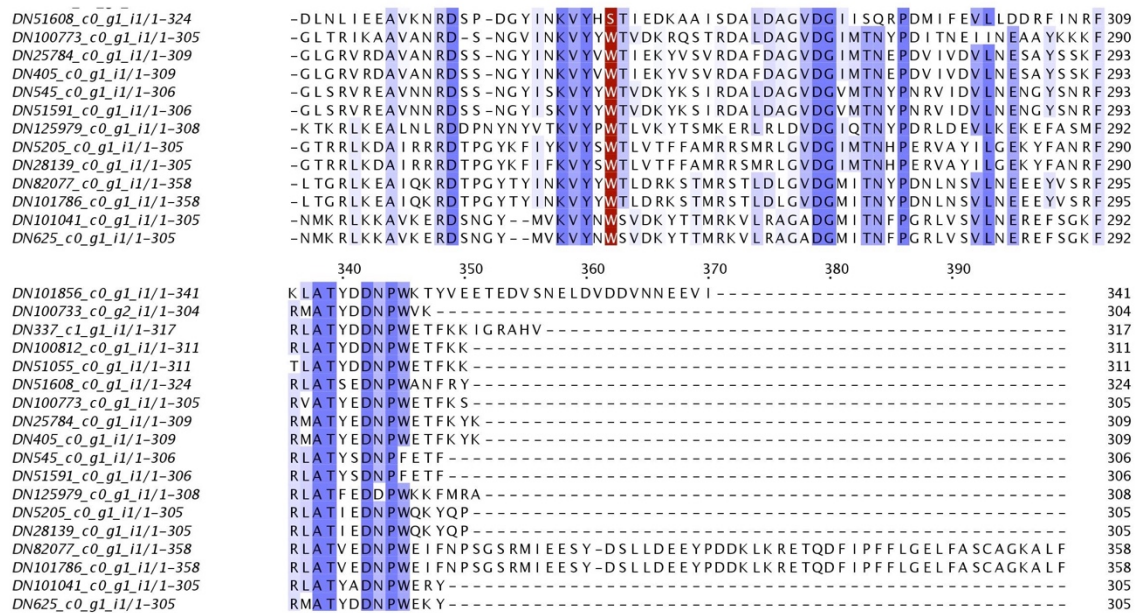


Figura 12 Alinhamento entre as sequências protéicas dos contigs do transcriptoma de *Loxosceles laeta*. Alinhamento com a ferramenta ClustalW com do programa Jalview, exibindo a similaridade entre as sequências dos contigs em tons de azul. Tons de azul mais escuros indicam um grau de conservação maior em relação aos tons mais claros. É possível observar resíduos de cisteínas conservados, que possivelmente formam quatro pontes dissulfeto e o resíduo de metionina localizado na alça hidrofóbica que ajuda a estabilizar o átomo de Mg²⁺. O sítio ativo formado pelo domínio H e W também está presente nas sequências marcado em vermelho

DN101856_c0_g1_i1/1-341
 DN28139_c0_g1_i1/1-305
 DN5205_c0_g1_i1/1-305
 DN125979_c0_g1_i1/1-308
 DN101786_c0_g1_i1/1-358
 DN82077_c0_g1_i1/1-358
 DN625_c0_g1_i1/1-305
 DN101041_c0_g1_i1/1-305
 gij 27372520|LOXLA/1-304
 gij 74814732|LOXLA/1-304
 DN100733_c0_g2_i1/1-304
 DN51055_c0_g1_i1/1-311
 gij 27372518|LOXLA/1-311
 DN100812_c0_g1_i1/1-311
 gij 27372516|LOXLA/1-311
 gij 41018058|LOXLA/1-311
 DN337_c1_g1_i1/1-317
 DN51608_c0_g1_i1/1-324
 DN100773_c0_g1_i1/1-305
 DN405_c0_g1_i1/1-309
 DN25784_c0_g1_i1/1-309
 DN51591_c0_g1_i1/1-306
 DN545_c0_g1_i1/1-306

DN101856_c0_g1_i1/1-341
 DN28139_c0_g1_i1/1-305
 DN5205_c0_g1_i1/1-305
 DN125979_c0_g1_i1/1-308
 DN101786_c0_g1_i1/1-358
 DN82077_c0_g1_i1/1-358
 DN625_c0_g1_i1/1-305
 DN101041_c0_g1_i1/1-305
 gij 27372520|LOXLA/1-304
 gij 74814732|LOXLA/1-304
 DN100733_c0_g2_i1/1-304
 DN51055_c0_g1_i1/1-311
 gij 27372518|LOXLA/1-311
 DN100812_c0_g1_i1/1-311
 gij 27372516|LOXLA/1-311
 gij 41018058|LOXLA/1-311
 DN337_c1_g1_i1/1-317
 DN51608_c0_g1_i1/1-324
 DN100773_c0_g1_i1/1-305
 DN405_c0_g1_i1/1-309
 DN25784_c0_g1_i1/1-309
 DN51591_c0_g1_i1/1-306
 DN545_c0_g1_i1/1-306

DN101856_c0_g1_i1/1-341
 DN28139_c0_g1_i1/1-305
 DN5205_c0_g1_i1/1-305
 DN125979_c0_g1_i1/1-308
 DN101786_c0_g1_i1/1-358
 DN82077_c0_g1_i1/1-358
 DN625_c0_g1_i1/1-305
 DN101041_c0_g1_i1/1-305
 gij 27372520|LOXLA/1-304
 gij 74814732|LOXLA/1-304
 DN100733_c0_g2_i1/1-304
 DN51055_c0_g1_i1/1-311
 gij 27372518|LOXLA/1-311
 DN100812_c0_g1_i1/1-311
 gij 27372516|LOXLA/1-311
 gij 41018058|LOXLA/1-311
 DN337_c1_g1_i1/1-317
 DN51608_c0_g1_i1/1-324
 DN100773_c0_g1_i1/1-305
 DN405_c0_g1_i1/1-309
 DN25784_c0_g1_i1/1-309
 DN51591_c0_g1_i1/1-306
 DN545_c0_g1_i1/1-306

DN101856_c0_g1_i1/1-341
 DN28139_c0_g1_i1/1-305
 DN5205_c0_g1_i1/1-305
 DN125979_c0_g1_i1/1-308
 DN101786_c0_g1_i1/1-358
 DN82077_c0_g1_i1/1-358
 DN625_c0_g1_i1/1-305
 DN101041_c0_g1_i1/1-305
 gij 27372520|LOXLA/1-304
 gij 74814732|LOXLA/1-304
 DN100733_c0_g2_i1/1-304
 DN51055_c0_g1_i1/1-311
 gij 27372518|LOXLA/1-311
 DN100812_c0_g1_i1/1-311
 gij 27372516|LOXLA/1-311
 gij 41018058|LOXLA/1-311
 DN337_c1_g1_i1/1-317
 DN51608_c0_g1_i1/1-324
 DN100773_c0_g1_i1/1-305
 DN405_c0_g1_i1/1-309
 DN25784_c0_g1_i1/1-309
 DN51591_c0_g1_i1/1-306
 DN545_c0_g1_i1/1-306

que possivelmente formam quatro pontes dissulfeto e o resíduo de metionina localizado na alça hidrofóbica que ajuda a estabilizar o átomo de Mg^{+2} . O sítio ativo formado pelo domínio H e W também está presente nas sequências marcado em vermelho

5.5.4 Visão geral da Filogenia da Aranha-marrom peruana

A princípio, foi realizado o alinhamento múltiplo que inferiu a similaridade com as enzimas de fosfolipases D encontradas na peçonha de *L. laeta* peruana contra outras proteínas do GenBank com melhores hits entre o alinhamento do Blastx. Com isso foi possível construir uma árvore filogenética utilizando o método de neighbor-joint disponibilizado pela plataforma interativa do programa Mega 7. Para dar suporte à árvore, foram utilizadas as 19 proteínas do transcriptoma, as proteínas obtidas com melhores hits como já mencionado e uma proteína externa de fosfolipase D do metabolismo humano. O teste estatístico de *Bootstrap* foi utilizado como ferramenta do Mega 7 para mensurar a confiabilidade dos ramos da árvore, foi utilizado o valor de 10 000 replicatas. O valor de confiabilidade considerado para esta análise foi 70.

A árvore filogenética vista na figura 14, estabelece uma relação entre os peptídeos alinhados e a similaridade dentre cada uma delas, ou seja, clusters de peptídeos que sejam mais semelhantes uns aos outros. Isso permite saber quais são as mais próximas entre si, mesmo entre proteínas homólogas, como é o caso e inferir uma possível classificação quanto a função ou nomenclatura. Conseguimos ver que todas as proteínas apresentam bastante proximidade. A árvore de similaridade obtida neste item (figura 14) apresentou suporte estatístico com média de bootstrap acima de 70% em geral. Isso significa que das várias combinações possíveis para o agrupamento das seqüências analisadas, a combinação observada em cada nóculo da árvore apresentada foi obtida em mais de 70% das réplicas aleatórias realizadas pelo software MEGA 7. Essa análise auxilia na determinação das possíveis funções específicas dessas fosfolipases. Consideramos a similaridade entre as sequências do transcriptoma de *L. laeta* peruana com o de outras aranhas como um indicador de supostos meios de ação para os peptídeos descritos. A árvore filogenética formada pelas fosfolipases-D presentes na peçonha de *L. laeta* apresentou coesão entre as proteínas do transcriptoma. As proteínas similares com espécies do gênero *Sicarius*, como esperado, apresentaram-se em um grupo peculiar.

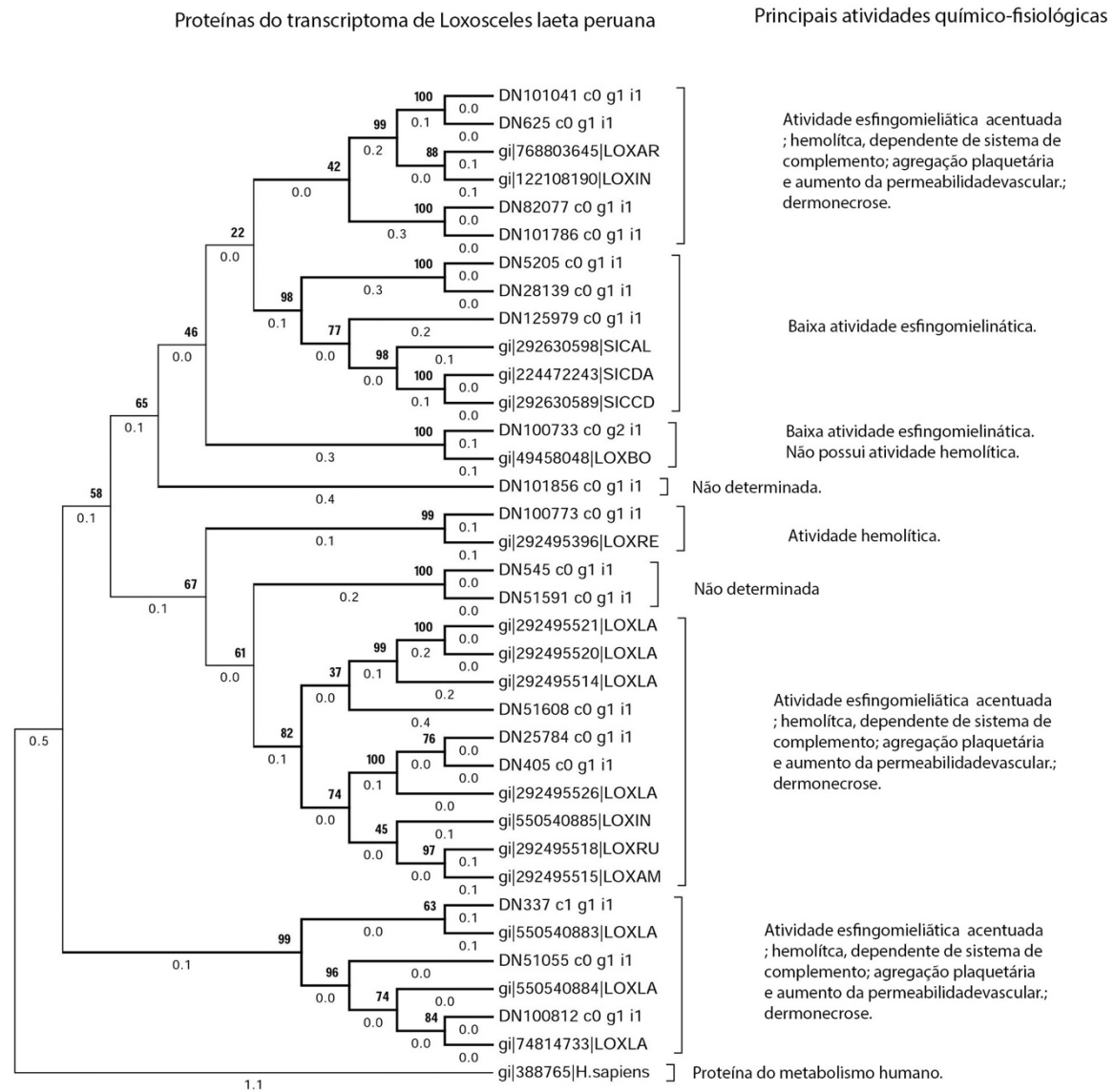


Figura 14 Cladograma do alinhamento múltiplo efetuado de similaridade com as enzimas de fosfolipases – D encontradas na peçonha de *L. laeta* contra outras proteínas do NCBI com melhores hits entre o alinhamento do Blastx. Os números colocados nos nódulos da árvore representam valores de bootstrap.

As proteínas obtidas do banco de dados não redundante do NCBI com alta similaridade com as proteínas do transcriptoma de *L. laeta* peruana, estão listadas com suas IDs seguidas de siglas referentes aos nomes das espécies. As proteínas sequenciadas foram separadas em 8 grupos de acordo com as proteínas mais parecidas e assim denominadas suas principais atividades funcionais no metabolismo humano. É possível verificar a presença de diversas isoformas de fosfolipases D com atividades específicas ou a serem estudadas.

Neste trabalho, além das toxinas já conhecidas, observamos a abundância da isoforma 1 de fosfolipase D na peçonha de *L. laeta* originada do Peru, o que pode ser importante para

uma análise mais apurada da filogenia das espécies, possibilitando avaliar a discrepância da expressão gênica. Este é um importante passo na compreensão da composição da peçonha e respectivos efeitos potenciais em humanos. As peçonhas das diferentes espécies de *Loxosceles* possuem características bioquímicas e farmacológicas parecidas, sendo a principal delas a atividade esfingomielinásica ou dermonecrótica.

6 CONCLUSÕES

- A análise bioinformática dos transcritos da glândula de peçonha da aranha *L. laeta* peruana identificou através da montagem de novo 143.237 transcritos.
- A anotação dos contigs identificados permitiu a identificação de 19 isoformas completas, no transcriptoma de *L. laeta* peruana, com elevada similaridade a PLDs.
- O estudo das proteínas completas identificadas permitiu a identificação de isoformas que apresentaram mutações em seus sítios ativos. Essas sequências tornam-se um alvo de estudos para seu uso na imunização para a produção de soro, visto que podem funcionar como epítomos eficientes para a geração de anticorpos neutralizantes.
- Através de um estudo detalhado percebemos a correlação entre as proteínas da peçonha de *L. laeta* peruana com as outras espécies de *Loxosceles* inclusive de *L. laeta* brasileira e de outras regiões, além de aranhas do gênero *Sicarius*.

7 PERSPECTIVAS

Pretendemos, em breve fazer a caracterização bioquímica e farmacológica de isoformas de PLDs de interesse, com objetivo de favorecer uma melhor compreensão dos mecanismos de ação do veneno e de cada uma de suas atividades enzimáticas. Dessa forma, as seqüências estudadas também poderão ser depositadas em bancos de dados, para que se tornem disponíveis à comunidade científica. A montagem do transcriptoma de glândulas de peçonha de *L. laeta* obtida contém muitos dados que futuramente podem ser revelados através de novas análises

voltadas para as demais classes de proteínas. A presença de componentes como neurotoxinas, fosfolipases, metaloproteases, proteínas ktx, alérgenos, hialuronidasas, serino-proteases, entre outras moléculas, anteriormente já descritas como abundantes em espécies do gênero, pode ser confirmada nos próximos estudos. Com a identificação de isoformas de interesse, pode-se obter novos alvos para o desenvolvimento de estudos envolvendo a produção de soros e de profilaxia.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ANSORGE, W.J. **Next-generation DNA sequencing techniques**. New Biotech. Vol. 25, 4:195-203, 2009.

APPEL, M. H. **Produção de ferramentas biológicas e estudo de proteínas dermonecróticas recombinantes de aranha marrom**. Curitiba: Universidade Federal do Paraná, 2006. 127 p. Tese (Doutorado). Programa de pós-graduação em biologia celular e molecular, Setor de ciências biológicas, Universidade Federal do Paraná, Curitiba, 2006.

ARAÚJO, S.C.; CASTANHEIRA, P. ALVARENGA, L.; MANGILI, O.C.; KALAPOTHAKIS, E.; CHÁVEZ-OLÓRTEGUI, C. **Protection against dermonecrotic and lethal activities of *L.intermedia* venom with a fused recombinant protein**. Toxicon, v. 41, p.261-267, 2003

AUBOURG S & ROUZE. **Genome annotation**. Plant Physiol. Biochem. 39: 181-193, 1991.

BARBARO, K.C.; CARDOSO, J.L.; EICKSTEDT, V.R.D.; MOTA, I. **Dermonecrotic and lethal components of *Loxosceles gaucho* spider venom**. Toxicon, v. 30, p. 331-338, 1992.

BINFORD, G.J., WELLS, M.A. **The phylogenetic distribution of sphingomyelinase D activity in venoms of Haplogygne spiders**. Comp. Biochem. Phys. B 135 (1), 25– 33, 2003.

BINFORD, G.J., CALLAHAN, M.S., BODNER, M.R., RYNERSON M.R., NUÑEZ P.B., ELLISON C., DUNCAN R. **Phylogenetic relationships of *Loxosceles* and *Sicarius* spiders are consistent with western Gondwanan vicariance**. Mol Phylog Evol. 2008.

BINFORD GJ, CORDES MHJ, WELLS MA. **Sphingomyelinase D from venoms of *Loxosceles* spiders: evolutionary insights from cDNA sequences and gene structure**. Toxicon. 45:547–560, 2005.

BONFIELD, J.K.; SMITH, K.F.; STADEN, R. **A new DNA sequence assembly program**. Nucleic Acids Res. v. 24, p. 4992-4999. 1995.

BORKAN J, GROSS E, LUBIN Y, ORYAN I. **An outbreak of venomous spider bites in a citrus grove**. Am J Trop Med Hyg. 52: 228– 230, 1995.

BRIGNOLI, P.M. **Beitrag zur Kenntnis der Scytodidae (Araneae)**. Rev. Suisse Zool. 83, 125–191, 1976.

BRIGNOLI, P.M. **Note sugli Scytodidae d'Italia e Malta (Araneae)**. Fragm. ent. 6, 121–166, 1979.

BRUSCA, R.C. & BRUSCA, G.J. **Invertebrados**. 7a. edição. Editora Guanabara Koogan. 2007.

CAMERON, H. D. **An etymological dictionary of North American spider genus names**. 274–330, 2005.

CARDOSO, J. L. C.; FRANÇA, F. O.; EICKSTEDT, V. D.; BORGES, I. & NOGUEIRA, M. T. **Loxoscelismo**: estudo de 242 casos (1980-1984). *Revista da Sociedade Brasileira de Toxicologia*. 1:58-60 1988.

CARDOSO, J. L. C. & DE CILLO, D. M. **Primeiro diagnóstico de “loxoscelismo” na casuística do Hospital Vital Brazil (HVB)**. *Memórias do Instituto Butantan*, 1950; 52(supl.):81.

CARDOSO, J. L. C. **Introdução ao estudo dos acidentes por animais peçonhentos**. In: *Plantas Venenosas e Animais Peçonhentos* (S. Schvartsman, org.). São Paulo: Sarvier. pp. 139-142, 1992^a.

CARDOSO, J. L. C. **Acidentes por Loxosceles (Loxoscelismo)**. In: *Plantas Venenosas e Animais Peçonhentos* (S. Schvartsman, org.). São Paulo: Sarvier. pp. 201- 204, 1992b.

CHAIM, O. M.; SADE, Y. B.; DA SILVEIRA, R. B.; TOMA, L.; KALAPOTHAKIS, E.; CHÁVEZ- OLÓRTEGUI, C. **Brown spider dermonecrotic toxin directly induces nephrotoxicity**. *Toxicology and Applied Pharmacology*, New York, v. 211, p. 64-77, 2006.

CHAVES-MOREIRA, D.; CHAIM, O. M.; SADE, Y. B.; PALUDO, K. S.; GREMSKI, L. H.; DONATTI, L.; DE MOURA, J.; MANGILI, O. C.; GREMSKI, W.; da SILVEIRA, R. B.; SENFF-RIBEIRO, A.; VEIGA, S. S. **Identification of a direct hemolytic effect dependent on the catalytic activity induced by phospholipase-D (dermonecrotic toxin) from brown spider venom**. *J. Cell. Biochem.*, v. 107, p. 655–666, 2009.

CHAVES-MOREIRA, D.; SOUZA, F. N.; FOGAÇA, R. T. H.; MANGILI, O. C.; GREMSKI, W.; SENFF-RIBEIRO, A.; CHAIM, O. M.; VEIGA, S.S. **The Relationship Between Calcium and the Metabolism of Plasma Membrane Phospholipids in Hemolysis Induced by Brown Spider Venom Phospholipase-D Toxin**. *J. Cell. Biochem.*, v. 112, p. 2529-2540, 2011.

CHAVEZ-OLORTEGUI, C; ZANETTI, V.C.; FERREIRA, A.P; MINOZZO, J.C.; MANGILI, O.C.; GUBERT, I.C. **ELISA for the detection of venom antigens in experimental and clinical envenoming by Loxosceles intermedia spiders**. *Toxicon*, 36 (1998), pp. 563–569

CHÁVEZ-OLÓRTEGUI, C., KALAPOTHAKIS, E., FERREIRA, A.M., FERREIRA, A.P., DINIZ, C.R.. **Neutralizing capacity of antibodies elicited by a non-toxic protein purified from the venom of the scorpion Tityus serrulatus**. *Toxicon* 35, 213–21, 1997.

COMPEAU PEC, PEVZNER PA, TESLER G. **How to apply de Bruijn graphs to genome assembly**. *Nature Biotechnology*. 29(11):987–991, 2011.

DA SILVA, P.H.; DA SILVEIRA, R.B.; APPEL, M.H.; MANGILI, O.C.; GREMSKI, W.; VEIGA, S.S. **Brown spiders and loxoscelism**. *Toxicon*, v. 44, n. 7, p. 693-709, 2004.

DA SILVA, P.H.; HASHIMOTO, Y.; DOS SANTOS, F.A.; MANGILI, O.C.; GREMSKI, W.; VEIGA, S.S. **Hematological cell findings in bone marrow and peripheral blood of rabbits after experimental acute exposure to Loxosceles intermedia (brown spider) venom**. *Toxicon*, v. 42, n. 2, p. 155-61, 2003.

DA SILVEIRA, R.B.; PIGOZZO, R.B.; CHAIM, O.M.; APPEL, M.H.; DREYFUSS, J.L.; TOMA, L.; MANGILI, O.C.; GREMSKI, W.; DIETRICH, C.P.; NADER, H.B.; VEIGA, S.S. **Molecular cloning and functional characterization of two isoforms of dermonecrotic toxin from *Loxosceles intermedia* (brown spider) venom gland.** *Biochimie*, v. 88, n. 9, p. 1241-53, 2006.

DAI, M.; THOMPSON, R. C.; MAHER, C.; et al. **NGSQC: cross-platform quality analysis pipeline for deep sequencing data.** *BMC genomics*, v. 11 Suppl 4, n. Suppl 4, p. S7, 2010.

de ANDRADE, S. A., MURAKAMI, M. T., CAVALCANTE, D. P., ARNI, R. K., TAMBOURGI, D. V. **Kinetic and mechanistic characterization of the Sphingomyelinases D from *Loxosceles intermedia* spider venom.** *Toxicon* 47: 380-386, 2006.

de ANDRADE, S. A., PEDROSA, M. F., de ANDRADE, R. M., OLIVA, M. L., VAN DER BERG, C. W., TAMBOURGI, D. V. **Conformational changes of *Loxosceles* venom sphingomyelinases monitored by circular dichroism.** *Biochem. Biophys. Res. Commun.* 327: 117-123, 2005.

de Roodt et al., 2002 A.R. de Roodt, O.D. Salomon, S.C. Lloveras, T.A. **Orduna Poisoning by spiders of *Loxosceles* genus.** *Medicina (B Aires)*, 62 (2002), pp. 83–94

DIAZ, J.H. **The global epidemiology, syndromic classification, management and prevention of spider bites.** *American Journal of Tropical Medicine and Hygiene*, v.71, n.2, p.239-250, 2004.

DURAI A. D. AND SCHULZ M.H. **Informed kmer selection for *de novo* transcriptome assembly** *Bioinformatics*. 32(11): 1670–1677, 2016.

Ewing, B. and Green, P. **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res.* 8: 186-194, 1998.

SCHMIEDER e EDWARDS, 2011 Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res.* 8: 175-185.

FELICORI ET AL., 2006 L. FELICORI, S.C. ARAUJO, R.A. DE AVILA, E.F. SANCHEZ, C. GRANIER, E. KALAPOTHAKIS, C. CHAVEZ-OLORTEGUI. **Functional characterization and epitope analysis of a recombinant dermonecrotic protein from *Loxosceles intermedia* spider** *Toxicon*, 48 (2006), pp. 509–519

FERNANDES-PEDROSA, M. F., JUNQUEIRA-DE-AZEVEDO I. L. M.; GONCALVES-DE-ANDRADE, R. M., VAN DEN BERG, C. W., RAMOS, C. R., HO, P. L.; TAMBOURGI, D. V. **Molecular cloning and expression of a functional dermonecrotic and haemolytic factor from *Loxosceles laeta* venom.** *Biochem Biophys Res Commun*, v. 298, p. 638- 645, 2002.

FERNANDES-PEDROSA, JUNQUEIRA-DE-ZEVEDO, I. L. M.; GONÇALVES- DE-ANDRADE, R. M.; KOBASHI, L. S.; ALMEIDA, D. D.; HO, P. I.; TAMBOURGI, D. V.; **Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags.** *Genomics*, v. 9, p. 1-12, 2008.

FERRER, V.P., DE MARI, T.L., GREMSKI, L.H., TREVISAN-SILVA, D., DA SILVEIRA, R.B., GREMSKI, W., CHAIM, O.M., SENFF-RIBEIRO, A., NADER, H.B., VEIGA, S.S. **A novel hyaluronidase from brown spider (*Loxosceles intermedia*) venom (Dietrich's Hyaluronidase): From cloning to functional characterization.** PLoS Negl. Trop. Dis. 7, e2206, 2013.

FERRER, V. P. **Clonagem e Expressão Heteróloga de Hialuronidase e Alérgeno presentes no veneno de aranha marrom (*Loxosceles intermedia*).** Biologia Celular e Molecular. Curitiba, Dissertação (Mestrado) - Universidade Federal do Paraná: 102, 2010.

FISCHER, M. L. **Comportamento sexual de *Loxosceles laeta* (Nicolet) (Araneae, Sicariidae): in uência da idade da fêmea.** Revista Brasileira de Zoologia, Curitiba, v. 24, n. 4, p. 865-872, 2007.

FOELIX, R.F. **Biology of Spiders.** Oxford University Press, Oxford, p. 419, 2011.

FOIL, L.D.; COONS, L.B.; NORMENT, B.R. **Ultrastructure of the venom gland of the brown recluse spider, *Loxosceles reclusa* Gertsch and Mulaik (Araneae:Loxoscelidae).** Int. J. Insect. Morphol. Embryol. 8: 325-334, 1979.

FUTRELL, J.M. **Loxoscelism.** Am J Med Sci, v. 304, n. 4, p. 261-7, 1992.

GAJARDO-TOBAR, R.R. **Mi experiencia sobre loxoscelismo.** Mem. Inst. Butantan. 33: 689-698, 1966.

GALIANO ME, HALL M. **Datos adicionales sobre el ciclo vital de *Loxosceles laeta* (Nicolet, 1849) (Araneae)** Physis, 32:277-288, 1973.

GERTSCH WJ. **The spider genus *Loxosceles* in South America (Araneae: Scytodidae).** Bulletin of American Museum of Natural History 136:119-183, 1967.

GERTSCH WJ, ENNIK F. **The spider genus *Loxosceles* in North America, Central America, and the West Indies (Araneae, Loxoscelidae).** Bulletin of American Museum of Natural History 175: 264-360, 1983.

GERSCHEMAN DE, P.B.S., SCHIAPELLI, R.D. **Caracteres morfológicos válidos en la sistemática del género *Sicarius* (Walckenaer 1847) Araneae: Sicariidae.** Acta Zool. Lilloana 35, 87-96, 1979.

GHOSH S, CHANCK **Analysis of RNA-Seq Data Using TopHat and Cufflinks.** Methods Mol Biol 1374: 339-361, 2016.

GONZAGA, M. O. ; SANTOS, A. J. ; JAPIASSÚ, H. F. **Ecologia e comportamento de aranhas.** Rio de Janeiro: Interciência, 2007. 400p.

GARBER, M.; GRABHERR, M. G.; GUTTMAN, M.; TRAPNELL, C. **Computational methods for transcriptome annotation and quantification using RNA-seq.** Nature methods, v. 8, n. 6, p. 469-77, 2011.

GRADA, A.; WEINBRECHT, K. **Next-generation sequencing: methodology and application.** The Journal of investigative dermatology, v. 133, n. 8, p. e11, 2013.

GREMSKI, L. H.; DA SILVEIRA, R. B.; CHAIM, O.M.; PROBST, C. M.; FERRER, V. P.; NOWATZKI, J.; WEINSCHUTZ, H.C.; MADEIRA, H.M.F.; GREMSKI, W.; NADER, H.B.; SENFF-RIBEIRO, A.; VEIGA, S.S. **A novel expression profile of the *Loxosceles intermedia* venomous gland revealed by transcriptome analysis.** Molecular Biosystems, v. 12, p. 2403-2416, 2010.

GUILHERME ET AL., 2001 P. GUILHERME, I. FERNANDES, K.C. BARBARO. **Neutralization of dermonecrotic and lethal activities and differences among 32±35 kDa toxins of medically important *Loxosceles* spider venoms in Brazil revealed by monoclonal antibodies.** Toxicon, 39 (2001), pp. 1333–1342

GUO, Y.; YE, F.; SHENG, Q.; CLARK, T.; SAMUELS, D. C. **Three-stage quality control strategies for DNA re-sequencing data.** Briefings in bioinformatics, 2013.
GRIMALDI D. AND ENGEL M.S. **Evolution of the Insects.** Cambridge University Press: New York, 2005.

GUIMARÃES G1, DIAS-LOPES C, DUARTE CG, FELICORI L, MACHADO DE AVILA RA, FIGUEIREDO LF, DE MOURA J, FALEIRO BT, BARRO J, FLORES K, SILVA W, TINTAYA B, YARLEQUE A, BONILLA C, KALAPOTHAKIS E, SALAS CE, CHÁVEZ-OLORTEGUI C. **Biochemical and immunological characteristics of Peruvian *Loxosceles laeta* spider venom: neutralization of its toxic effects by anti-loxoscelic antivenoms.** Toxicon. 70:90, 2013

HAAS BJ ET AL. **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** Nat Protoc 8:1494–1512, 2013.

HAGEN, J. B. **The origins of bioinformatics.** Nature reviews. Genetics, v. 1, n. 3, p. 231–6, 2000.

HOGAN, BARBARO & WINKEL. **Loxoscelism: Old obstacles, new directions** Annals of emergency medicine, January 2005

JÄNES J. et al. A comparative study of RNA-seq analysis strategies. Brief Bioinform., 16, 932–940, 2015.

JEON, Y.-S.; PARK, S.-C.; LIM, J.; CHUN, J.; KIM, B.-S. **Improved pipeline for reducing erroneous identification by 16S rRNA sequences using the Illumina MiSeq platform.** Journal of Microbiology, v. 53, n. 1, p. 60–69, 2015.

KALAPOTHAKIS, E.; PENAFORTE, C. L.; BEIRÃO, P. S.; ROMANO-SILVA, M. A.; CRUZ, J. S.; PRADO, M. A.; GUIMARÃES, P. E.; GOMEZ, M. V.; PRADO, V. F. **Cloning of cDNA's coding neurotoxic peptides from the spider *Phoneutria nigriventer*.** Toxicon, v. 36, p. 679-685, 1998.

KALAPOTHAKIS, E., ARAUJO, S. C., DE CASTRO, C. S., MENDES, T. M., GOMEZ, M. V., MANGILI, O. C., GUBERT, I. C. & CHAVEZ- OLORTEGUI, C. **Molecular cloning,**

expression and immunological properties of LiD1, a protein from the dermonecrotic family of *Loxosceles intermedia* spider venom. *Toxicon*, v. 40, p. 1691-1699, 2002.

KALAPOTHAKIS, E., CHATZAKI, M., GONCALVES-DORNELAS, H., DE CASTRO, C. S., SILVESTRE, F. G., LABORNE, F. V., DE MOURA, J. F., VEIGA, S. S., CHAVEZ-OLORTEGUI, C., GRANIER, C. and BARBARO, K. C. **The Loxtox protein family in *Loxosceles intermedia* (Mello-Leitao) venom.** *Toxicon*, v. 50, p. 938-946, 2007.

KASABOV N. **Evolving connectionist systems:** the knowledge engineering approach, second edn. Springer, New York, Secaucus, NJ, USA. 2007.

KATZ, Y.; WANG, E. T.; AIROLDI, E. M.; BURGE, C. B. **Analysis and design of RNA sequencing experiments for identifying isoform regulation.** *Nature methods*, v. 7, n. 12, p. 1009–15, 2010.

KATZ, Y.; WANG, E. T.; SILTERRA, J.; et al. **Sashimi plots** : Quantitative visualization of alternative isoform expression from RNA-seq data. , , n. 1, 2013.

KIM, D.; PERTEA, G.; TRAPNELL, C.; et al. **TopHat2:** accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, v. 14, n. 4, p. R36, 2013.

KLEINER, D. J.; STETLER-STEVENSON, S. W. **Structural biochemistry and activation of matrix metalloproteases.** *Current Opinion in Cell Biology*, London, v. 5, p. 891-897, 1993.

KUSMA, J., CHAIM O.M., WILLE, A.C., FERRER, V.P., SADE, Y.B., DONATTI, L., GREMSKI, W., MANGILI, O.C., VEIGA, S.S. **Nephrotoxicity caused by brown spider venom phospholipase-D (dermonecrotic toxin) depends on catalytic activity.** *Biochimie* 90: 1722-1736, 2008.

LEE S, LYNCH KR. **Brown recluse spider (*Loxosceles reclusa*) venom phospholipase D (PLD) generates lysophosphatidic acid (LPA).** *Biochem J* 391: 317-323, 2005.

LISTER R, PELIZZOLA M, DOWEN RH, HAWKINS RD, HON G, TONTI-FILIPPINI J, NERY JR, LEE L, YE Z, NGO QM, EDSALL L, ANTOSIEWICZ-BOURGET J, STEWART R, RUOTTI V, MILLAR AH, THOMSON JA, REN B, ECKER JR. **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 462(7271):315-22, 2009.

LUCAS, S. **Spiders in Brazil.** *Toxicon*, v. 26, p.759-772, 1998.

MACCHIAVELLO, Z. - **Cutaneous arachnoidism or gan grenous spot of Chile.** *Puerto Rico J. publ. Hlth.* 22: 425 - 466, 1947.

MACHADO, L. F., LAUGESSEN, S., BOTELHO, E. D., RICART, C. A., FONTES, W., BARBARO, K. C., ROEPSTORFF, P. & SOUSA, M. V. **Proteome analysis of brown spider venom:** identification of loxnecrogin isoforms in *Loxosceles gaucho* venom. *Proteomics*, v. 5, p. 2167-2176, 2005.

MALAUQUE ET AL., C.M. MALAUQUE, J.E. CASTRO-VALENCIA, J.L.C. CARDOSO, F.O.S. FRANÇA, K.C. BARBARO, H.W. **Fan Clinical and epidemiological features of definitive and presumed loxoscelism in São Paulo, Brazil** *Revista do Instituto de Medicina trop.*, 44. pp. 139–143, 2002

MARKOVIC-HOUSLEY, Z.; MIGLIERINI, G.; SOLDATOVA, L.; RIZKALLAH, P.J.; MULLER, U.; SCHIRMER, T. **Crystal structure of hyaluronidase, a major allergen of bee venom.** *Structure* 2000, 8, 1025–1035.

MANRIQUEZ AND SILVA, 2009. J.J. MANRIQUEZ, S. SILVA **Cutaneous and visceral loxoscelism: a systematic review.** *Revista Chilena de Infectologia*, 26 (2009), pp. 420–432

MARCHLER-BAUER, A., BRYANT, S.H. **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res.* 32, 327–331, 2004.

MARCHLER-BAUER A, LU S, ANDERSON JB, CHITSAZ F, DERBYSHIRE MK, DEWEESE-SCOTT C, FONG JH, GEER LY, GEER RC, GONZALES NR, GWADZ M, HURWITZ DI, JACKSON JD, KE Z, LANCZYCKI CJ, LU F, MARCHLER GH, MULLOKANDOV M, OMELCHENKO MV, ROBERTSON CL, SONG JS, THANKI N, YAMASHITA RA, ZHANG D, ZHANG N, ZHENG C, BRYANT SH. **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic Acids Res.* 2011.

MARCHLER-BAUER, A.; ZHENG, C.; CHITSAZ, F.; DERBYSHIRE, M. K.; GEER, L. Y.; GEER, R. C.; GONZALES, N. R.; GWADZ, M.; HURWITZ, D. I.; LANCZYCKI, C. J.; LU, F.; LU, S.; MARCHLER, G. H.; SONG, J. S.; THANKI, N.; YAMASHITA, R. A.; ZHANG, D.; BRYANT, S. H. "CDD: Conserved domains and protein three-dimensional structure". *Nucleic Acids Research*. 41 (Database issue): D348–D352, 2012

MARDIS ER. **The impact of next-generation sequencing technology on genetics.** *Trends Genet.* 3:133–14, 2008.

MARDIS ER. **New strategies and emerging technologies for massively parallel sequencing: applications in medical research.** *Genome Med*; p 1:40, 2009.

MARDIS ER. **A decade's perspective on DNA sequencing technology.** *Nature*; p 470:198–203, 2011.

MARQUES DA SILVA, E.; FISCHER, L. M. **Distribuição das espécies do gênero *Loxosceles Hei-necken & Lowe, 1835 (Araneae; Sicariidae)* no Estado do Paraná.** *Revista da Sociedade Brasileira de Medicina Tropical, Uberaba*, v. 38, p. 331–335, 2005.

MARQUES-DA-SILVA, E.; SOUZA-SANTOS, R.; FISCHER, M. L.; RUBIO, G. B. G. ***Loxosceles* spider bites in the state of Paraná, Brazil: 1993–2000.** *Journal of Venomous Animals and Toxins incling Tropical Diseases, Botucatu*, v. 12, n. 1, p. 110–123, 2006.

MARTIN-EAUCLAIRE, M.-F., ALAMI, M., GIAMARCHI, A., MISSIMILLI, V., ROSSO, J.-P., BOUGIS, P.E. **A natural anatoxin, Amm VIII, induces neutralizing antibodies against the potent scorpion alpha-toxins.** *Vaccine* 24, 1990–6. doi:10.1016/j.vaccine.2005.11.016, 2016.

MENDES, T.M., OLIVEIRA, D., FIGUEIREDO, L.F.M., MACHADO-DE-AVILA, R. A., DUARTE, C.G., DIAS-LOPES, C., GUIMARÃES, G., FELICORI, L., MINOZZO, J.C., CHÁVEZ-OLORTEGUI, C. **Generation and characterization of a recombinant chimeric protein (rCpLi) consisting of B-cell epitopes of a dermonecrotic protein from *Loxosceles intermedia* spider venom.** *Vaccine* 31, 2749–2755. doi:10.1016/j.vaccine.2013.03.048. 2013.

MOREIRA-FERREIRA, A.M., KALAPOTHAKIS, E., DINIZ, C.R., CHÁVEZ-OLÓRTEGUI, C., **In vivo protection against *Tityus serrulatus* scorpion toxins by immunization of mice with a non-toxic protein.** *Toxicon* 36, 333–9, 1998.

METZKER, M. L. **Sequencing technologies - the next generation.** *Nature Reviews Genetics*, v. 11, n. 1, p. 31-46, Jan. 2010.

MINISTÉRIO DA SAÚDE, 2012. **Ministério da Saúde Sistema nacional de notificação de agravos. Ministério da Saúde, Brasília, Distrito Federal. 2012**

MINISTÉRIO DA SAÚDE, 2016. **Ministério da Saúde Sistema nacional de notificação de agravos.** Ministério da Saúde, Brasília, Distrito Federal. 2016

MION G., GOYFFON M. **Les envenimations graves.** Arnette édit., Rueil-Malmaison, 2000

NAGARAJAN, N., et al. **Finishing genomes with limited resources: lessons from an ensemble of microbial genomes,** *BMC Genomics*; p 11, 242, 2010.

NAGARAJAN, N. AND POP, M. **Sequence assembly demystified,** *Nat. Rev. Genet.*, 14, 157-167, 2013.

NEULANDS, G., 1975. **A revision of the spider genus *Loxosceles* Heineken & Lowe, 1835 (Araneae: Scytodidae) in Southern Africa with notes on the natural history and morphology.** *J. Entomol. Soc. S. Afr.* 38 (2), 141–154.

NEULANDS, G., **A new spelaeon species of *Loxosceles* (Araneae: Sicariidae).** *J. Entomol. Soc. S. Afr.* 43 (2), 367–369, 1980.

OLIVEIRA, K.C.; ANDRADE, G1UST1, A L ; SILVA, WD.; TAMBOURGI, D,V **Sex-linked variation of *Loxosceles intermedia* spider venoms.** *Toxicon*, v.37, p. 217- 221, 1999.

PABINGER, S.; DANDER, A.; FISCHER, M.; et al. **A survey of tools for variant analysis of next-generation genome sequencing data.** *Briefings in bioinformatics*, v. 15, n. 2, p. 256–78, 2014.

PALUDO, K. S.; BISCAIA, S. M.; CHAIM, O. M.; OTUKI, M. F.; NALIWAIKO, K.; DOMBROWSKI, P. A.; FRANCO, C. R.; VEIGA, S. S. **Inflammatory events induced by brown spider venom and its recombinant dermonecrotic toxin: a pharmacological investigation.** *Comp Biochem Physiol C Toxicol Pharmacol*, v. 149, p. 323-333, 2009.

PATEL, K.D., MODUR, V., ZIMMERMAN, G.A. **The necrotic venom of the brown recluse spider induces dysregulated endothelial cell- dependent neutrophil activation: differential induction of GM-CSF, IL-8, and E-selectin expression.** *J. Clin. Invest* 94, 631–642, 1994.

PANAFTOSA. **Consulta técnica sobre accidentes con animales ponzoñosos en Latinoamérica informe final de la consulta técnica sobre accidentes con animales ponzoñosos en latino América.** São Paulo, Brasil. 2007.

PAULI, I.; PUKA, J.; GABERT, I.C.; MINOZZO, J.C. **The efficacy of antivenom in Loxoscelism treatment.** *Toxicon*, v.48, n.2, p.123-137, 2006.

PAVLOPOULOS, G. A; OULAS, A.; IACUCCI, E.; et al. **Unraveling genomic variation from next generation sequencing data.** *BioData mining*, v. 6, n. 1, p. 13, 2013.

PEVZNER, P.A., TANG, H., AND WATERMAN, M.S. **A new approach to fragment assembly in DNA sequencing.** In: RECOMB 01, pp. 256–267, 2001.

PHADTARE, S., AND INOUYE, M. **Genome-wide transcriptional analysis of the cold shock response in wild-type and cold-sensitive, quadruple-csp-deletion strains of Escherichia coli.** *J Bacteriol* 186, 7007–7014, 2004.

PINTO, A.C., MELO-BARBOSA, H.P., MIYOSHI, A., SILVA, A., AND AZEVEDO, V. **Application of RNA-seq to reveal the transcript profile in bacteria.** *Genet Mol Res* 10, 1707–1718, 2011.

PLATNICK, N. I. The world spider catalog, version 13.5. American Museum of Natural History, online at <http://research.amnh.org/iz/spiders/catalog>. Acesso em 12/05/2016.

RAMOS-CERRILLO, B. et al. **Genetic and enzymatic characterization of sphingomyelinase D isoforms from the North American fiddleback spiders *Loxosceles boneti* and *Loxosceles reclusa*.** *Toxicon*, 44, 507–514, 2004.

REES, R.S.; O'LEARY, P.; KING, L.E. **The pathogenesis of systemic loxoscelism following brown recluse spider bites.** *Journal of Surgical Research*, v.35, p.1-10, 1983.

REES, R.S.; NANNEY, L.B.; YATES, E.; KING, L.E. **Interaction of brown recluse spider venom on cell membranes: the inciting mechanism?** *Journal of Investigational. Dermatology*, New York, v.83, p.270-275, 1984.

RIBEIRO, L. A.; EICKSTEDT, V. R. D.; RUBIO, G. B. G.; KONALSAISEN, J. F.; HANDAR, Z.; ENTRES, M.; CAMPOS, V. A. & JORGE, M. T. **Epidemiologia dos acidentes por aranhas do gênero *Loxosceles* Heinecken & Lowe no Estado do Paraná (Brasil).** *Memórias do Instituto Butantan*, 1993;55:19-26. 47, 1993.

SAMS, H. H.; DUNNICK, C. A.; SMITH, M. L.; KING, L. E. **Necrotic arachnidism.** *Journal of American Academy of Dermatology*, Saint Louis, v. 44, n. 4, p. 603–608, 2001.

SANABRIA, H., ZAVALETA, A., **Panorama epidemiológico del loxoscelismo en el Perú.** *Rev Peru Med Exp. Salud Publica* 14 (2), 33–41, 1997.

SANTOS, V.L.P.; FRANCO, C.R.C.; VIGGIANO, R.L.L.; SILVEIRA, R.B.; CANTÃO, M.P.; MANGILI, O.C.; VEIGA, S.S.; GREMSKI, W. **Structural and ultrastructural description of the venom gland of *Loxosceles intermedia* (brown spider).** *Toxicon*. 38: 265-285, 2000.

SANTOS, A.J.; BRESCOVIT, A.D.; JAPYASSÚ, H. F. **Diversidade de aranhas: sistêmica, ecologia e inventários de fauna.** In: GONZAGA, Marcelo. O. et al. (Org.). *Ecologia e comportamento de aranhas.* Rio de Janeiro: Editora Interciência, 2007.

SHENDURE, J.; JI, H. **Next-generation DNA sequencing.** *Nature biotechnology*, v.26, n.10, p.1135-1145, 2008.

SCHENONE, H. **Toxic pictures produced spiders bites in Chile: latrodectism and loxoscelism.** *Rev Med Chil*, v. 131, n. 4, p. 437-44, 2003.

SCHENONE, H.; SAAVEDRA, T.; ROJAS, A.; VILLARROEL, F. **Loxoscelism in Chile.** Epidemiologic, clinical and experimental studies. *Rev Inst Med Trop Sao Paulo*, v. 31, n. 6, p. 403-15, 1989.

SCHMIEDER, R., AND EDWARDS, R. **Quality control and preprocessing of metagenomic datasets.** *Bioinformatics*, 2011.

SCHULZ MH, ZERBINO DR, VINGRON M, BIRNEY E. **Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels.** *Bioinformatics*. 2012;28:1086–1092.

SEZERINO, U.M.; ZANNIN, M.; COELHO, L.K.; GONCALVES JUNIOR, J.; GRANDO, M.; MATTOSINHO, S.G.; CARDOSO, J.L.; VON EICKSTEDT, V.R.; FRANCA, F.O.; BARBARO, K.C.; FAN, H.W. **A clinical and epidemiological study of Loxosceles spider envenoming in Santa Catarina, Brazil.** *Trans R Soc Trop Med Hyg*, v. 92, n. 5, p. 546-8, 1998.

SILVEIRA, A. L. **Primeiro registro sinantrópico de Loxosceles laeta (Nicolet, 1849) (Araneae, Sicariidae) no Município do Rio de Janeiro, Estado do Rio de Janeiro.** *Rev. Soc. Bras. Med. Trop.* [online]. vol.42, n.6, pp.723-726. ISSN 0037-8682, 2009.

SILVESTRE, F.G., DE CASTRO, C.S., DE MOURA, J.F., GIUSTA, M.S., DE MARIA, M., ALVARES, E.S., LOBATO, F.C., ASSIS, R.A., GONÇALVES, L.A., GUBERT, I.C., CHÁVEZ-OLÓRTEGUI, C., KALAPOTHAKIS, E. **Characterization of the venom from the brazilian brown spider Loxosceles similis Moenkhaus, 1898 (Araneae, Sicariidae).** *Toxicon* 46, 927–936, 2005.

SMITH, C.W. ; MICKS, D.W. **The role of polymorphonuclear leukocytes in the venom of the brown recluse spider, Loxosceles reclusa** *Lab Invest*, 22, pp. 90–93, 1970.

STRAUSBERG, R.L.; LEVY, S.; ROGERS, Y.H. **Emerging DNA sequencing technologies for human genomic medicine.** *Drug Discov Today*, v.13, n. 13/14, p. 569-77, 2008.

SURGET-GROBA, Y.; MONTOYA-BURGOS, J. I. **Optimization of de novo transcriptome assembly from next-generation sequencing data.** *Genome Research*, v. 20, n. 10, p. 1432–1440, 2010.

TAMBOURGI, D.V., GONÇALVES-DE-ANDRADE, R.M., VAN DEN BERG, C.W. **Loxoscelism: From basic research to the proposal of new therapies.** *Toxicon* 56: 1113-1119, 2010.

TAMBOURGI DV, MAGNOLI FC, VAN DEN BERG CW, MORGAN BP, DE ARAUJO PS, ALVES EW, DA SILVA WD. **Sphingomyelinases in the venom of the spider *Loxosceles intermedia* are responsible for both dermonecrosis and complement-dependent hemolysis.** *Biochem Biophys Res Commun.* 251:366–373, 1998.

TRAPNELL, C.; PACHTER, L.; SALZBERG, S. L. **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics (Oxford, England),* v. 25, n. 9, p. 1105–11, 2009.

TRAPNELL, C.; ROBERTS, A.; GOFF, L.; et al. **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nature protocols,* v. 7, n. 3, p. 562–78, 2012a.

TRAPNELL, C.; ROBERTS, A.; GOFF, L.; et al. **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nature protocols,* v. 7, n. 3, p. 562–78, 2012b.

TREVISAN-SILVA, D.; GREMSKI, L. H.; CHAIM, O. M.; DA SILVEIRA, R. B.; MEISSNER, O. M.; MANGILI, O. C.; BARBARO, K. C.; GREMSKI, W.; VEIGA, S. S.; SENFF-RIBEIRO, A. **Astacin-like metalloproteases are a gene family of toxins present in the venom of different species of the brown spider (genus *Loxosceles*).** *Biochimie,* v.92, p.21-32, 2010.

Trinity online manual. <http://trinityrnaseq.sourceforge.net>.

VAN ASWEGEN G, VAN ROOYEN JM, VAN DER NEST DG, VELDMAN FJ, DE VILLIERS TH, OBERHOLZER G. **Venom of a six-eyed crab spider, *Sicarius testaceus* (Purcell, 1908), causes necrotic and haemorrhagic lesions in the rabbit.** *Toxicon.* 1997;35:1149–1152, 1997.

VEIGA, S.S., DA SILVEIRA, R.B., DREYFUS, J.L., HAOACH, J., PEREIRA, A.M., MANGILI, O.C., GREMSKI, W., **Identification of high molecular weight serine-proteases in *Loxosceles intermedia* (brown spider) venom.** *Toxicon* 38, 825–839, 2000a.

VEIGA, S.S., FEITOSA, L., DOS SANTOS, V.L., DE SOUZA, G.A., RIBEIRO, A.S., MANGILI, O.C., PORCIONATTO, M.A., NADER, H.B., DIETRICH, C.P., BRENTANI, R.R., GREMSKI, W., **Effect of brown spider venom on basement membrane structures.** *Histochem. J.* 32, 397–408, 2000b.

VEIGA, S.S., ZANETTI, V.C., FRANCO, C.R.C., TRINDADE, E.S., PORCIONATTO, M.A., MANGILI, O.C., GREMSKI, W., DIETRICH, C.P., NADER, H.B., **In vivo and in vitro cytotoxicity of brown spider venom for blood vessel endothelial cells.** *Throm. Res.* 102, 229–237, 2001a.

VEIGA, S.S., ZANETTI, V.C., BRAZ, A., MANGILI, O.C., GREMSKI, W., **Extracellular matrix molecules as targets for brown spider venom toxins.** *Braz. J. Med. Biol. Res.* 34, 843–850, 2001b.

VETTER, R.S. & VISSHER, P.K. **Bites and stings of medically important venomous arthropods.** *Int Journal Dermatol,* v.37, 9.481-496, 1998.

VUITIKA, L., GREMSKI, L.H., BELISÁRIO-FERRARI, M.R., CHAVES-MOREIRA, D., FERRER, V.P., SENFF-RIBEIRO, A., CHAIM, O.M., VEIGA, S.S. **Brown spider phospholipase-D containing a conservative mutation (D233E) in the catalytic site:** Identification and functional characterization. *J. Cell. Biochem.* 114, 2479–2492, 2013.

XIA, X.H., XIE, Z., SALEMI, M., CHEN, L., WANG, Y. **An index of substitution saturation and its application.** *Mol. Phylogenet. Evol.* 26 (1), 1–7, 2003.

WANG, J.F. **Two new species of spiders of the genus *Loxosceles* from China.** *J. Hebei Norm. Univ. Suppl.*, 13–15, 1994.

WASSERMAN GS & ANDERSON PC. **Loxoscelism and necrotic arachnidism.** *Journal of Toxicology. Clinical Toxicology*, 21: 451-472, 1984.

WATERHOUSE, A. **Jalview Version 2:** A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191, 2009.

ZERBINO, D.R. and BIRNEY, E. **Velvet:** Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Reserch.* 2008. 18: 821–829.

ZERBINO, D.R. **Using the Velvet de novo assembler for short-read sequencing technologies.** *Curr. Protoc. Bioinformatics* 31. 11.5.1–11.5.12, 2010.

ZUPUNSKI, V., KORDIS, D., GUBENSEK, F., **Adaptive evolution in the snake venom Kunitz/BPTI protein family.** *FEBS Lett.* 547, 131–136, 2003.

9 ANEXOS

9.1 Anexo 1 – Fasta com as sequências do cDNA de *L. Laeta* correspondentes aos contigs obtidos através da montagem de novo de RNA-Seq.

>DN405_c0_g1_i1

```
TTTTTTTTTTCACACTAGAGATCAATGCTTTAATAGTCAAAAATAATTACAGTTTTGCGGC
ATATAATAGAACTTCTGTAAAGGTCAAGAAAACGAATATACTTGTTTGTGAGTTTAAAAG
TTCAATACTGAAGTGTCTTCTCATAAAAAGCCTTATGTTAAATTAAGGTTATTATCTTT
GTTCAAGTTTTACAGTTATAGTTTTCAATACGAAATATGATTCTGATTTCTGAACAG
CATCTAAAGACGTTGTCTATTGGACTCCTCCACCAAGTATATCGTCATAAGAAGTGCTTG
TGTAATTGTTTATTTGATTTAAAAGTTTCCCACGGGTTGTCTTCATATGTAGCCATTCT
GAACTTGCTGCTGTAAGCGCTTTCATTTAAGACATCAACAATAACGTCTGGCTCGTTAGT
CATGATACCGTCTACCCAGCGTCAAAGGCGTCTCTGACTGAAACGTATTTTTCTATCGT
CCAGACGTACACTTTGTTGATGTAACCGTTTGAAGAGTCTCGATTTGCCACGGCATCTCT
GACTCGACCAAGACCTCGACCAATACAGTTGGTGATACCATCGCTTTGCCACACGTGCTC
TCTATCCTTTACTCCGCCTTTATTTAAAGCATTACGAGTCGCGCTCAGGTCGTCATTTCC
GGAAAAGTCGTAGCCTACTTTGTCAAGAAGGTCTGCATGGTCAGAAATCTTAAAAGCCTC
TCTGAATCCTGATATGAATGCATAATGATCAATATCTGGGATAGATAATACTATGTATGC
TCTTCCACCGTTATTACCTGCATTCCAATAGTGGATAAACAGCTTCTGTGCGAACATATT
TCCCGCTTTGTAAGCCGTTGAGGAAGAAAGATCGCCTGTTTTCAAATCGAACACAACATAA
GACCAGTTCTGGGTGATACTTGGAAATTACCAGGGGTCGTAGCTCGTTGCAATCCGTGCAA
AAAGTCGTTAACATTTTCCCACTTGGTACACCACCTTTTGCAGTCACAGGGCACCCCGTG
GTAAGTGTACTCTGCTATAGCGTCATCATCAAAGGTAATATCCGTTTCGATTGAATTCCG
GCCAAGGTCCACGAATTCGTCAATTTGATAGATCTCGTTTACCATATGCCCATGTTCCA
TATTGGCCTGCGGTTGTCTGCTCGTTCGATCACCTCAGTTTCCGCACCCTGCAGCAGGAC
GGCACAGAAACCCAGCAGTAAAGAAATCTAACAGCATATCGAAAGTGCTCTTCCCTTCT
TTCAGATGCAGTTCATTCAGTACTTTATATATCATTCTCTATATGTCAAACCTAGAGGTA
AAAGTTACGCTTACACATGTGAAAGATGTGGCTGTCATACAATGTAACGGGTGTTTTGC
AGCTTTCCCTTGCTCATCTGTCATAATTCTTCTGGAACATGTGTTTGTGGGAAGCTTG
AGTGGTATAAAACAAGGAG
```

>DN51055_c0_g1_i1

```
TTTTTTTTTTTTTTGTTTCTGCATACATTTTATTAAGCACAAATATCATATTTTGAACAAG
ATTTTTGAATAATACAGAATGCTGCATCTTGATGCATCTGCAAAGAAAAGAGTATATAAG
AATTCTGCATCACACAAGCTTCCAGACGTTTTGTTTTATATAAACATTTTCCAGCCATG
TGAAGTGAACACTACAGTATTTGTTGAAGCTTGTATTTTTTAAAAGTCTCCCATGGATTG
TCGTCGTAAGTTGCCAACGTGTAGTTATCCTTGATCCATCTTCGTTGAGGACGTCGATA
ATAACGTACGGAAAATTGGTCATTATCCATCAACGCCAACATCTAGTGCTGTTCTCGTT
GTTGAATATTTGTCTACAGACCAGTAGTAACTTTGCTAATGAATCCATTTTCCGAGTCC
CTTCTTTTTACAATTTCTTTCAGTCTAGCCATGTCACCAAGTGCGGCCCAATTGGTAAGG
CCATCACTTAGCCAGACGTGACCATCCACTCCAGCTTTTTTAAAGGCTTCGTGAACAGTAA
TCCAGTGAGGGTAGGCTTGGCAAGTATGGGCCAGATAAGTCGTATCCTACTTTGTCCAAC
AAATTCTCATGACCTTCTGCCTTAAAGTACTTCTTAAATGTTTCGTATAAATTCGTAATGC
GCGATATCAGGCAATGACAATACTACGTATGCTTCCGCCATCGTTACCATTGTTCCAG
TAATTCTGTAGAAGTTCCTTTGCTACATTTTCTCCGGCCGGACGTAATTGATCGTTGCTC
AGGCTACCGGTCTTTAAGTCCAAAACAAACAGTATGAACCCATCACGATACTTGGCATT
CCTGGCGTTGTGATTTCTCAGTGTTTTCAGAAATACATTGAAATACTCCCATCTGATG
CAGTCCCTACCGAAGTCGCAAGGCGTTCCATGGTAACTGTAGGTGGGCACCGCTCCCTTA
AAAGTAATGTCCGCTTCTATTGCGTTTGCACCATCATTTAAAAGGTTGGTATCTGTTTT
ACAGCGTTTACCATGTGGCCCAGGTTCCAAATTGGACGACGTTATCAGCTCGTTCCACCA
ACATCCGTTACGGCTCCCTGTAGGATTACGGTCCAGCATCCTAAGATTAACGCTAAGTGA
ACGTACATAATGAAATGCGAGTTCAGAGCTAGCACGACAATTGATGCACGGACTCAGTGA
CTTATATAGGCCTTTTATCGATTAATATGAG
```

>DN337_c1_g1_i1

```
ATACATCGGTGACTGGAGTTAGACGTGTGCTCTTCCGATCTTTTTAAAAGTCTCCCATGG
ATTGTCGTGTAAGTTGCCAACCTGTATTTGTCTTGTATCCATCTTCGTTGAGGACGTC
AATAATAACGTACGGATAATTGGTCATTATCCATCAACGCCAACATCGAGTGCTGCCTT
```

CGTTGTTGATACTTTGTCTACAGACCAGTAGTAAATTTTATTGATAAATCCATTTGCCGA
 ATCCCTGCTTTTAAATAGCTTCTTTAGTCGATCAATGTTTAAAAGTGGCGCAAAATTGGT
 GAGGCCATCACTTTGCCAGACGTGACCCTCCACTCCGGCTTTTTTAAAGGCTTCGTGAGT
 TTGATCCAGTGTGGGCAGGCTTGGCACGTATGGGCCAGAGAAGTCGTATCCTACTTTCTC
 CAACAAGTTCTCATGACCTTCTCTTAAGTACTTCTTTAAATGCTCTTATGAATTCGTA
 ATGCGCGATATCAGGTAATGACAACACTACGTACGCTCTCCACCATTATTACCATTGTT
 CCAGTAATGCTCTAGAAGTTGCTTTGCTACATTTTCTCCGGCTTTTCTACTTCATCGTT
 GTTCAAGCTTCCCGTCTTCAAGTCCAAAACAAAACAGTATAAACTGCTCACGATAGTTGGC
 ATTTCTGTTGTTGTGAATCTCTCAGTGTGGAAGAAATACGTTGAAATACTCCCATCT
 GATGCAGTCCCTACCGAAGTCGCAAGGCGTTCCGTGGTAAGTGTAGGTAGGCACCGATCC
 CTTAAAAGTAACATCCGCTTCTAATGCGTTTGCACCAAGATCCAAAAGGTTGGTATCTG
 TTTTACAGCGTTCACCATATGAGCCAGGTTCCAAATTGGACGACGTTATCAGCTCGTTC
 ACCAACATCCGTTACGGCTCCCTGTAGGATTACGGTCCAGCATCCTAAGATTAACGCTAA
 GTGAACGTACATAATGAAATGCGAGAGATCGGAAGAGCGTCTGTGA

>DN100812_c0_g1_i1

CTCATATTAATCGATAAAAAGGCCTATATAAGTCACTGAGTCCGTGCCTCAATTGTCGTGC
 TAGCTCTGAACCTCGCATTTCATTATGTACGTTCACTTAGCGTTAATCTTAGGATGCTGGA
 CCGTAGTCTACAGGGAGCCGAAACGGATGTTGGTGAACGAGCTGATAACCGTCTGCCAA
 TTTGGAACTTGGGCCACATGGTAAACGATGTTAAACAGATAACCACTTTTTAAATGATG
 GTGCAAACGCAATAGAAGCGGACATTACTTTTAAAGGGAGCGGTGCCACCTACAGTTACC
 ATGGAACGCCTTGGACTTCGGTAGGGACTGCATCAGATGGGAGTATTTGACGTATTTT
 TGCAAACACTGAGAGATTACACAACGCCAGGAAATCCAAGTATCATGAGAAGTTTATAC
 TTTTTTTTTGGACTTGAAGACGGGTAGCTTGAACAACAATGAAGTGAAGAAAGCCGGGG
 AAAATGTAGCAAAGGGACTTCTACAGAATTAAGTGAACAATGGTAACAATGGCGGAAGAG
 CATACGTAGTATTGTCATTGCCTGATATCGCGCATTACGAATTTATACGAACATTTAAAG
 AAGTACTTAAGGCAGAAGGTCATGAGAATTTGTTGGACAAAGTAGGATACGACTTATCTG
 GCCATACTTGGCAAAGCCTACCCTCACTGGATTACAGTTACGAAAGCCTTTAAAAAAGCTG
 GAGTGGATGGTCACTGCTGGCTAAGTGTGCTTACCAATTGGGCGCCACTTGGTGACA
 TGGCTAGACTGAAAGAAATTGTAAGAAAGGGACTCGGAAAATGGATTCATTAGCAAAG
 TTTACTACTGGTCTGTAGACAAATATCAACAACGAGAACAGCACTAGATGTTGGCGTTG
 ATGGAATAATGACCAATTTCCGTACGTTATTATCGACGTCCTCAACGAAGATGGATACA
 AGGATAACTACAGGTTGGCAACTTACGACGACAATCCATGGGAGACTTTTAAAAAATAAC
 AAGCTTCAACAAACTGTAGTTGCACCTTACATGGCTGAAAATGTTTATATATAAAAA
 AAACCTGGAAGCTGTGTGATGCAAAATCTTATATACTCTTGTTTACAGATTCTACA
 CGATGCAGTATCCCTATTATTAATAAATCTTGTTCAAAAATATTTGTGCTTAATAAAAT
 TTAATCAGAAGC

>DN100773_c0_g1_i1

CGACTCTAGCAGATGCTCAAGCAATACGTGTCCATGTGGAGAAGGTGTATATATTATATG
 TTGGAGAAGGTATATACAGGTATTTCGTAGACAGATTGCTTGCACATTACGGTTGCTCC
 CTTCCATCACCATGTTGCCGTATATTGCTTTAGTTTTAGGGTGTGGGGCGTGTATCCG
 AGAGTGCTCAAACAGATGTTGGTGAACGAGCAGACAAACGTAGGCCTGTATGGATCATGG
 GCCACATGGTCAACGATCCCCACAGATAGATGAGTTCGTGAACCTTGGGGTGAATTCCA
 TCGAAACGGACGTGCTTTTCGATGACAACGCCAATCCTGAATACACATATCACGGTATTC
 CATGCGATTGCGGAAGAAGTTGCGGGCATTATGAGTATTTGACGTTTTTCTAAAAAATC
 TCGAATCCGCACAACACCTGGTCATTCAAAGTACCATGAAAACTTGTCTTGGTTGTGT
 TTGATTTGAAAACCTGGTAGCCTCTACGATAATCAAGCAAACGATGCCGGAAAATAAG
 CAAAAAGTCTGTTACAGAATACTGGAACAATGGCAATAATGGTGGAAAGCATACTTG
 TGTTATCCATACCCCATCTAAAACATTATCAACTGATAACAGGATTTCAAGAAACACTTA
 AAACTGAGGGACATGAAGAGCTGTTGGATAAAGTTGGATATGACTTCTCTGGAAACGACG
 AAATCAGTGACGTTGAGAATGCTTACAAAAAAGCCGGTGTAAACGGGTCATGTGTGGCAGA
 GCGACGGTATCACTAACTGTTAACGCGTGGTCTTACTCGAATAAAGGCAGCTGTGGCAA
 ACAGAGACTCAAACGGAGTCATTAACAAGGTGTAATACTACTGGACTGTGGACAAGCGTCAAT
 CGACCAGAGATGCACCTGATGCTGGAGTTGACGGCATAATGACCAATTATCCTGATATTA
 CTAATGAAATCATCAATGAAGCCGTTACAAGAAAGAAATTCAGAGTTGCCACATACGAAG
 ACAACCCTTGGGAGACATTCAAGAGCTAGATTATCTAGGTTGATTCTGTGCAACACATTA
 CAATCCGGATTTTTGGTTTTCCATAGAATTTGTTGAATGACCAATTTGCTGTACAGAATT
 GAAAAATTATACTGTGTCAAACGGTTGTTTAAATTTTTATTTTCGGTTCCGAAAAATTA
 AATGTTATAATTTTTTCAATTTATGAAAAAGTAAATTTTTTCAGTAATAAAATTTTCAGCTATG
 CAAAAA

>DN51591_c0_g1_i1

AAAAACCCAAAGAGTTTCATACTTTGATTTGGATAGTCACACGAACTTCTTCACTGCAACA
 TGATATTCAATATTGTTTTGACTATAGGATTTTTCAGCCGTTTTATTAGAGGGAGCCGAAA
 CTGATGTCAGCGAACGCGCAGACAATCGCCGGCCAATATGGAATATGGGTCACATGGTGA

ATGCTATTTACCAAATCGACGAGTTCGTGGATCTCGGAGCAAATTCCTAGAAAATAGACA
 TAACTTTTTGATAAGAAGGCCAAAGCGGATTACACTTACCACAAGATACCATGTGACTGTT
 TCAGAACATGTAGAAAGTGGGAATACGTCAATGACTACTTGAAAGCAGTGCGACGAGCTA
 CAACGCCTGGTTATTCAAAATATCGTTCTCGACTAGTGCTACTTGTGTTTGACTTGAAAA
 CAACCAAGCTTACCACGAATACGGCATACAAAGCTGGCGAAGACTTTGCAAGGAGGCTTC
 TGGAGCACTACTGGAACAACGGCAATGACGGCGGAAGGGCTTATATAGTGTATCAATCC
 CTGATGTTAATCATTATACGTTTATCAACGGATTTGGAGAAGCCCTAAGAAGCAGTGGTC
 ATCCGGAGTTAATGCAGAAAGTTGGTTACGATTTCTCTGGAATGACGACCTAAGCACAA
 TACGCAACGCATTGCGCAAAGCCGGTGTAAATGACAAGGAACACATCTGGCAAAGCGATG
 GCATACCAATTGCTTAGCACGTGGTCTTAGTCGAGTAAGAGAGGCTGTGAATAACAGAG
 ACTCCTCAAATGGATACATCAGTAAAGTCTACTACTGGACGGTTGACAAAATACAAGTCTA
 TAAGAGATGCACCTTGATGCTGGTGTGATGGTGAATGACTAACTATCCGAACCGTGTTA
 TCGATGTCCTGTAACAAAACGGTTACAGCAAGATTGAGATTGGCTACCTACAGCGACA
 ACCCTTTCGAAAACCTTCTAGAATTGATGTGTTTTTCCGCCTACTGTAACCTCAATATCC
 ATATGTATGAACGGTAGTAAAAAAGTTTCAGAAACATAAGGCAATGGACATTATGAATGA
 AAATTTCCCAAAGCAACAATGATGCTCGTTGGATATTATAGCAGCGTTCTCAAGGAACT
 CAGTGTCTAATTTTGTGAATTTCAATTTATGAGTTTTTCGAAAACGCTTTTGAATTCTAC
 TGGATCCTTCTGTAATTTAACTTTATTTTATTTCTTTCTTTGGGCAATATTTAGCAT
 CATTTGTATGAGAAAATAATTTTCTACAAATAGTTTACAATGAGAAAATTTATGTTAGGTG
 ACAAGTGTCAAAAACGCAATGTTTCGCTTTAGTTATTGTTGAAACATTTTCTATAAAAAT
 TGAAATAAAAGATCGTAAATATTATTGTGTCAGTAATAAAATTTCTCAGATATGAAAAAA

>DN545_c0_g1_i1

TATAATATCCAGTGTGTTATTCCATGTTTCATTTAATCATTTAATGTTAAGATTTGTTTT
 GTTTCAACCAGATGCTACATTTTCACTGGTTGATACTTTTTTACCCGAAGAAAAGCTACA
 GGGTACTGCAGTGGTTTACTGCAGTGGTATACTTCTCTCGATTTGCAATCGGTAGAAATT
 CCTTCATATTCTTGAAGATCGCTCTTTGCCACAATCTATGTAAATTTCTATAAAACATTT
 TAGTTAATTTTTTCGATTATGACCTTTTGGTGTGCCAGCTGATGTTGCTTCCATTGATTG
 ATTGATTGTGGTGTTTTCACTATCTCTCGTAATAGATACAAATATACGTGAAAGTCGAA
 AACTATGAATCATACTGAGAAATTTTATTACTGACAATAATTTTACGATCTTTTATTT
 CACATTTTATAGAAAATGTTTCAACAATAACTAAAGCGAACATTGCGTTTTATGACACTT
 GTCACCTAACATAAAATTTTCTCATTGTAACCTATTTGTAGAAAATTTTCTCATACAAA
 TGATGTAATAATTTGCCAAAGAGAAAAGAAAATAAAATTTTAAATTACAGAAGGATC
 CAGTAGAATTCAAAAGCGTTTTTCGAAAACCTATAAAATTGAAATTCACAAAATTAGAACAC
 TGAGTTCCTTGAGAACGCTGCTATAATATCCAACGAGCATCATTGTTGCTTTGGGAAAT
 TTTTATTACATAATGTCCATTGCCTTATGTTTCTGAAAATTTTTTACTACCGTTCATACAT
 ATGGATATTGAGAGTTACAGTAGGCGGAAAACACATCAATTCTAGAAGGTTTCGAAAGG
 GTTGTGCGCTGTAGGTAGCCAACTGAATCTGTTGCTGTAACCGTTTTTCGTTCCAGGACATC
 GATAACACGGTTCCGATAGTTAGTCATTACACCATCAACACCAGCATCAAGTGCATCTCT
 TATAGACTGTATTTGTCAACCGTCCAGTAGTACTTTACTGATGTATCCATTTGAGGA
 GTCTCTGTTATTCACGCCTCTTTACTCGACTAAGACCACGTGCTAAGCAATTTGGTGAT
 GCCATCGCTTTGCCAGATGTGTTCCCTTGTCAATTTACACCGGCTTTGCGCAATGCGTTGCG
 TATTGTGCTTAGGTCGTCAATTTCCAGAGAAATCGTAACCTACTTTCTGCATTAACCTCCAG
 ATGACCACTGCTTCTTAGGGCTTCTCCAAATCCATTGATAAACGTATAATGATTAACATC
 AGGGATTGACAACACTATATAAGCCCTTCTCCGTCATTGCCGTTGTTCCAGTAGTGCTC
 CAGAAGCCTCCTTGCAAAGTCTTCGCCAGCTTTGTATGCCGTATTTCGTGGTAAGCTTGGT
 TGTTTTCAAGTCAAACACAAGTAGCACTAGTCGAGAACGATATTTTGAATAACCAGGCGT
 TGTAGTCTGTCGACTGCTTTCAAGTAGTCACTATTGACGTATTTCCACTTTTCTACATGTTT
 GAAACAGTACATGGTATCTTGTGGTAAGTGAATCCGCTTTGGCCTTCTTATCAAAAAGT
 TATGTCTATTTCTAGGGAATTTGCTCCGAGATCCACGAACTCGTCGATTTGGTAAATAGC
 ATTCACCATGTGACCCATATTTCCATATTGGCCGGCGATTGTCTGCGCGTTCGCTGACATC
 AGTTTCGGCTCCCTCTAATAAAACGGCTAAAAATCCTATAGTCAAAAACAATTTGAATAT
 CATGTTGCAGTGAAGAAGTTTCGTGTGACTATCCAATCAAAGTATGAACTCTTTGGGTTT
 TT

>DN25784_c0_g1_i1

ATGAGCACAAGGGAAAGCTGCAAAACACCCGTTACATTGTATGACAGCCACATCTTTTAC
 ATGTGTGAAGCGTAACTTTTACCTCTAGTTTACATCTAGAGAAATGATATATAAAGTAC
 TGAATGAACTGCATCTGAAAGAAGGAAGAGCACTTTGATATGCTGTTAGATATTTCTTT
 ACTGCTGGGTTTTCTGTGCCGCTCTGTCAGGGTGCAGGAACTGAGGTGATCGAACGAGC
 AGACAACCGCAGCCAAATATGGAACATGGGGCATATGGTAAACGAGATCTATCAAATTTGA
 CGAATTCGTGGACCTTTGGCGCAATTCGAACGGAATTTACCTTTGATGATGACGC
 TATAGCAGAGTACACTTACCACGGGGTGCCTGTGACTGCAAAAGGTGGTGTACCAAGTG
 GGAAAATGTTAACGACTTTTTGCACGGATTGCAACGAGCTACGACCCCTGGTAAATTTCAA
 GTATCACCCAGAAGTGGTCTTAGTTGTGTTTCGATTTGAAAACAGGCGATCTTTCTTCTC

AACGGCTTACAAAGCGGAAATATGTTCCGACAGAAAGCTGTTTATCCACTATTGGAATGC
 AGGTAATAACGGTGAAGAGCATACATAGTATTATCTATCCCAGATATTGATCATTATGC
 ATTCATATCAGGATTCAGAGAGGCTTTTAAAGAATTCTGACCATGCAGACCTTCTTGACAA
 AGTAGGCTACGACTTTTCCGGAAATGACGACCTGAGCGCGACTCGTAATGCTTTAAATAA
 AGGCGGAGTAAAGGATAGAGAGCACGTGTGGCAAAGCGATGGTATCACCAACTGTATTGG
 TCGAGGTCTTGGTCGAGTCAGAGATGCCGTGGCAAATCGAGACTCTTCAAACGGTTACAT
 CAACAAAGTGTACGTCTGGACGATAGAAAAATACGTTTCAGTCAGAGACGCCTTTGACGC
 TGGGGTAGACGGTATCATGACTAACGAGCCAGACGTTATTGTTGATGTCTTAAATGAAAG
 CGCTTACAGCAGCAAGTTCAGAATGGCTACATATGAAGACAACCCGTGGGAAACTTTTAA
 ATACAAATAAACAAATTACACAAGCACTTCTTATGACGATATACTTGGTGGAGGAGTCCAA
 TAGACAACGTCTTTAGATGCTGTTCAAGAAATCAGAATCATATTTGATTTGAAAACTA
 TAACTGTGAAAACTTGAACAAAGATAATAACCTTAATTTAACATAAAGGCTTTTTATGAGA
 AGAACACTTTAGTATTGAACTTTTAACTCACAAACAAGTATATTCGTTTTCTTGACCTT
 TACAGAAGTTCTATTATGCCGCAAACTGTAATTTTTGACTATTAAGCATTGATC
 TCTAGTGTGAAAAA

>DN100733_c0_g2_i1

TCAGCTGGTGGCAGCTGTTGATGTCAATCGTACTCTGATGTCATTTCAAATGATATAA
 TGGGATGAAGAGCTATATAAAGATCAGGTATAATCACAGAAGGACACTTGCCTCCTAGTT
 TCCTCTGCTTACCATGTGGCTCTCCGCTGTCATATCCTTTATAGGATTTGCTGTCTTCC
 TCCAGGAAGCAAATGGGTACGTTGAAGAAAGGGCTGACAGTCGGAAGCCAATATGGGATA
 TAGCGCACATGGTAAACGATCTTGCCTGGTGGACGAATATTTAGGCGATGGCGCAAACG
 CTCTTGAGGCCGATCTAGCCTTTAAATCTGACGGTACTGCAGAGAAAATGTACCATGGAA
 CACCATGTGACTGCTTTAGAAGCTGTTCCGCAACAGAGAAATTCGACAAATACATGGACT
 ACATACGTGAAATTACAACACCTGGCAGCAGCAAGTTCAAATCTGGAATGTTATTGCTAA
 TAATAGACCTTAACTGAATGGCATAGAACCAAATGTCGCGTATGCAGCCGGAAAAAGCA
 CTGCAGAGACACTTTTGAAGAATTAAGTGGCAGAACGGTAAAAGTAAAGCCAGGGCATATA
 TTGTGCTTTCTTTGAGACCATAACCAGACCAGATTTTATAAGCGGTTTCAAAGACGCGA
 TTGATGCGAGTGGACACACGGAATATGAAAAAATAGGCTGGGATTTCTCCGGCAACG
 AGGACTTGGGGCAGATCCGACAGAGCTACCAGAAATATGGGATCGATGATCACATTTGGC
 AAGGCGACGGAATCACTAACTGTCTGCCACGAGGTGACTATCGTCTGACAGAAGCCATGA
 AAAAGAAAAATGATCCCAGTACAAGTACACCAAGAAAGTGTACACCTGGAGCATTGATA
 AAGAACTTCCATTAGAAATGCACTAAGACTTGGAGTGGATGCTGTGATGACTAACTATC
 CAGCAGTGTGAAGAGCGTTTTGCAGGAAAGTGAATTTCTTGGCAGTCATAGAATGGCAA
 CGTATGATGATAATCCATGGGTGAAATAAAAACTGCTTCCGGATCGACTGGCTAAATGGA
 AGTGTGAAACCGATTGTTGAAATCAAATTTCTTCTGTAACCTGGAATAATTGAGACCT
 CACTTTTCCGCTTTGTTCTTTAATCTGTTTATGTTTTTTTTACATCAGCATTTCAATAAG
 GTTGTAGTCAATAATAAAGATTTTTTTTTTTTT

>DN100733_c0_g1_i1

TTTTTTTTTTTTTCAAACAAAAATCTTTATTATCTACTATCAACGCTCATTGAAGTGCT
 GATGTAAAAAACAATAAACAGATTAAGAACAAGCGGAAAAGTGAAGTCTCAATTA
 TTCCAGGTTACACGAAGAAATTTGATTTCAACAATCGGTTTCAACACTTCCATTTAGCCA
 GTCGATCCGGAAGCAGTTTTTATTTACCCATGGATTATCATCATACGTTGCCATTCTAT
 GACTACCGGAGAATTCATTTCTGCAAGACGCTCTTACACGCTGCTGGATAGTTAGTCA
 TGACAGCATCCACTCCAAGTCTTAGTGCATTTCTAATGGAAGATTCTTTATCAATGCTCC
 AGGTGTACACTTTCTGGTGTACTTGTAGTCCGGATCATTTTTCTTTTTCATGGCTCTG
 TCAGACGATAGTCACCTCGTGGCAGACAGTTAGTGATTCCGTCGCCTTGCCAAATGTGAT
 CATCGATCCCATATTTCTGGTAGGCTCTGCGGATGTCGCCAAGTCTCGTTGCCGGAGA
 AATCCCAGCCTATTTTTTTCATATAGTTCCGTGTGTCCACTCGCATCAATCGCGTCTTTGA
 AACCGCTTATGAAATCTGGTCTGGTTATGGTCTCAAGAGAAAGCACAATATATGCCCTGG
 CTTTACTTTACCGTTCTGCCAGTAATTTCAAAGTGTCTCTGCAGTGCTTTTTCCGG
 CTGCATACGGCAGATTTGGTTCTATGCCATTCAGTTAAGGTCTATTATTAGCAATAACA
 TTCCAGATTTGAACCTGCTGCTGCCAGGTGTTGTAATTTACAGTATGTAGTCCATTGATT
 TGTCGAATTTCTGTTCCGCAACAGCTTCTAAAGCAGTCACATGGTGTTCATGTGATA
 TTTTCTCTGCAGTACCGTCCAGATTTAAAGGCTAGATCGGCCTCAAGAGCGTTTGCGCCAT
 CGCCTAAATATTCGTCCACCAGCGCAAGATCGTTTACCATGTGCGCTATATCCCATATTG
 GTTCCGACTGTCAGCCCTTTCTTCAACGTACCCATTTGCTTCTGGAGGAAGACAGCAA
 ATCCTATAAAGGATATGACAGCGGAGAGCCACATGGTGAAGCAGAGGAAACTAGGAGGCA
 AGTGTCTTCTGTGATT

>DN101786_c0_g1_i1

GTTTCAGTTCATCTAAGATGCTGCTCAGTCTCATTCTCGTCTGTCTGATAAGGCCATCTAT
 ACTTTTAGATGCAGCTAATGGAGAAGGTTTCATGGAGCGTGTCTGTTGATGGACGACGACC

AATATGGAACATCGCACACATGACCAACGATATCGAAATTGTGAGAACATATTTGAACGA
 TGGTGCTAATGCTCTCGAGTTTGACATAGAATTCAACAAAAACGGAAAACCTGTGAGGAC
 GTTTCATGGTTCTCCCTGTGACTGTTGTAGAAAATGTGACAGGACTACAACATTTTCTAC
 ATTCATTGAGCAAATGAGGAAATTCCTACGCTGGTGTATCCGGAATTTAAAAAGCAACT
 GGTGTTACTCATGCTGGATTTGAAAGTGACGGATTTTGATAAAAAGTCTGGCTTACACAGC
 TGGACAGAACGTAGCTGAGGTGCTTCTGGATAACTTCTGGCAAAGAGGTAAAAGTAAGCC
 ACAAGCCTACCTTCTGCTATCAATTCATCTCTCGACTACGCTAAACTTATTAAGGTTT
 TAATGATAAGCTGAAAGATGAAGGATTTGAGGTTTATCAGGAAAAAGTGGGTGTTGACTT
 CTCTAATAACGATGACTTAGACGATGTTTCGTAACCTTGAAGAGTATAGGATTTGAGAG
 TCGCATTGGCAAAGTATGGTATTACAACTGCCTTCCCCGTCTCACCGGCCGTCTAAA
 GGAGGCTATACAGAAAAGGATACTCCTGGATACACGTATATCAATAAGGTATATTACTG
 GACCTTGGATCGAAAGTCTACTATGAGATCTACGTTAGACTGGGAGTCGACGGTATGAT
 CACTAATTACCCTGATAACTTGAACAGCGTGAATGAAGAAGAATACGTAAGTCGGTT
 CAGATTTGGTACTGTAGAAGACAACCCATGGGAAATATTCAATCCTAGTGGATCACGTAT
 GATCGAAGAAAGCTATGATAGTCTGCTAGATGAAGAATATCCTGATGACAAATTAAGG
 AGAGACGCAGGATTTTATACCATTTTCTGGGTGAGTTGTTGCTTCTGTGCCGGCAA
 AGCATTATTTAATGCAAAAACACAGGCTATTTTCAGTTCATGGATCAATCTTTTTTCTA
 TGAAGACATTGCTTCAAGTACTGCGCTATGTACCACAGAATGAGAAACCACTTTGAAATG
 AGAAATCATGAGCAGATACTTAGCAATAAATCGTAAAAAATAAAGCAAGCAAGGTA
 AATAAAAATAAAAATATTTCCGTTCACTTTTTCGAAATATTAATCTGGCAGTGTCA
 TCGTTTGGTGGCGCCATCTGCAACCGAACTTAGGCAGAACTGGTGGTTGTGCTTCTCCA
 CGAAAGCGCTGATAACCGTCAAATGCGAGAAGCCCAACAACAATGGTCCCGAAAACACCA
 AACAACGCAGCAACTATCCTAGCTGCCTGTCCATATCGACATGTGGGTATGGCAACGAAC
 ATCAGAATGGCAGCAACCAGGTATAAGATACTAAAAATGACATCGTTGATGAAAATCGAA
 AGGGAGAAATTGATGATAGTGCAGAGCTTGTCCAACGAAATGGTGTACAAAACCAAATG
 ATTATGGCCAAAATGAGAGTGGTTACGACAACGAAATTAAGAAGCTGTACTTTGCATGG
 TCTGCACAGTACGCGCCTTGGCTTTCGCTGATTATCCACAAGACCCACACATGCTTCC
 ATAATCCTAAGAAATCCCTGGCCTGCTCTTGATGACTCCGTGCTGAGGGCCAAACCTCCC
 GAGCTGCCTTGGGCGGCAGCCATCACTGCTCACTCCTTCAACGCAACAGGTCCTTCGCTA
 ACTTCTCCGCCTTCTCAACTGCCCTGACACAAGCTCTCTGAACGCAACAGCATTTCTCT
 TCCTT

>DN82077_c0_g1_i1

GTTTTTTTTTTTTTTTCCATGCCAAAATATTTATTTTATTTTATTTACCTTGCTTGC
 TTATATTTTTTTACGATTTATTGCTAAGTATATCTGCTCATGATTTCTCATTGAAAAGTG
 GTTCTCATTCTGTGGTACATAGCGCAGTACTTGAAGCAATGTCTTCATAGAAAAAAGAT
 TGATCCATGAACTGAAAATAGCCTGTGTTTTGCATTAATAATGCTTTGCCGGCACAA
 GAAGCAAACAACCTCACCCAAGAAAAATGGTATAAAATCCTGCGTCTCTTTTTTAATTTG
 TCATCAGGATATTCTTCATCTAGCAGACTATCATAGCTTTCTTCGATCATACGTGATCCA
 CTAGGATTGAATATTTCCCATGGTTGTCTTCTACAGTAGCCAATCTGAACCGACTTACG
 TATTCTTCTTCAATTCAGCAGCTGTTCAAGTTATCAGGGTAATTAGTGATCATACCGTCCG
 ACTCCCAAGTCTAACGTAGATCTCATAGTAGACTTTTCGATCCAAGGTCCAGTAATATACC
 TTATTGATATACGTGTATCCAGGAGTATCCCTTTTCTGTATAGCCTCCTTTAGACGGCCG
 GTGAGACGGGAAGGCAGTTTGAATACCATCACTTTGCCAAATGCGACTCTCAAATCCT
 ATACTCTTCAAGGTTTTACGAACATCGTCTAAGTCATCGTTATTAGAGAAAGTCAACACCC
 ACTTTTTCTGATAAACCTCAAATCCTTCATCTTTCAGCTTATCATTAAAACCTTTAATA
 AGTTTAGCTAGTCCGAGAGATGGAATTGATAGCAGAAGGTAGGCTTGTGGCTTACTTTTA
 CCTTTTGGCAGAAGTTATCCAGAAGCACCTCAGCTACGTTCTGTCCAGCTGTGTGAAGCC
 AGACTTTTATCAAAATCCGTCACTTTCAAATCCAGCATGAGTAACACCAGTTGCTTTTTA
 AATTCCGGATCACCAGGCGTAGTGAATTTCTCATTGCTCAATGAATGTAGAAAATGTT
 GTAGTCTGTACATTTTCTACAACAGTACAGGGAGAACCATGAAACGTCCTCACAGGT
 TTTCCGTTTTTGTGAATTCTATGTCAAACCTCGAGAGCATTAGCACCATCGTTCAAATAT
 GTTCTCACAATTTGATATCGTTGGTCAATGTGTGCGATGTTCCATATTGGTGTGCTGCA
 TCAACAGCACGCTCCATGAAACCTTCTCCATTAGCTGCATCTAAAAGTATAGATGGCCTT
 ATCAGACAGACGAGAATGAGACTGAGCAGCATCTTAGATGAACTGAAC

>DN28139_c0_g1_i1

GTCGATATTTAATCAGGGCTAAAAAATGACTAGTTTTCTTCGTATCAGATCATGACGTAAG
 ACAATGGATCACTTTGCTATGAAAATATACAAATGCACAAACATTTTATACTATATGA
 AATTATTCTACATTTTGCAAAAAATGACATCTCCCTTTTATCATTGTTCCAATGGTAGGAT
 CAGGGCTGATACTTCTGCCACGGGTTGTCTTCAATAGTAGCCAGTCTGAAACGATTAGCA
 AAATACTTTTCTCCAAGAATATATGCTACACGTTCCGGATGATTGGTCAATTATACCATCT
 ACACCTAATCTCATTGATCTTCTCATTGCAAAAACGTGACCAAAGTCCAGGAATATACT
 TTGAAAATGAATTTGTACCCAGGGGTGCTCTTCTACGTATGGCATCTTTCAGTCGCTA

GTACCTCGTGGGAAACAGCTGGTAATCCATCAGCTTGCCAAATGTTACTGGTGATGTTCC
 ATTCTCTTCAGCATCCTACGAGTTGCTTCAAGGTCATCATTACCAGTGAAATTTATCCCA
 AACTTATCATTATATCTTTTCATAACCTTCTCTTTCGAATAATTGTATGTTGAAAACCGTTG
 ATGAAATCGAAATGACGGATACTTGGGATGTTGAGCAATATGTATGCTCTTGCAGTACCA
 TTTCTTTTTGCCAGTAATGATCAATGAGTTTATTTGCTATATCTTACCAGCAGCGTAG
 GCTTTGCTTAGATATTTCGTTGAAGCTTTAGATCAAGTGCAAGTAATATCAGTTGATCA
 CGAAATCTTTGCTATCTGGTGTAGTAACCAGCCGGATGTGATCCAAATACTTGACGAAG
 TCTTCTTTCTGGTGCATATTCGGCCGCGAGTCGCAGGGCACACCATGGTGTGTTTGTTC
 GCTATTCCATTGTAATTGAACACGATGTGCAATTCAATAGCGTTAGCTCCTCTTTTCAGA
 AAAGGGTTGATCTGGTCAATTGCGTTTACCATATGTGCAATGTTCCAAAATGGACGTTTA
 GAATCAGCGAATACATATTTTTGCGTCCCCTGCAATACGATGGATAATGCTACTAAGTAT
 GCAGTGGCACTAACATATCGAAGTAAATCTGCAGATTCCCTTGTGTGATGAAATTCAG
 AACATAAATATTATCTGTAACAAAACGGTAAGCAGAAGTTAG

>DN5205_c0_g1_i1

GTTCCCTTAGCAGGCTCCGGAGCTCAAATTCTGAAGAATTCCTGGTCCGAAAGAGATAAT
 ATTTATGTTCTGAATTTTCATCGACACAAGGGAATCTGCAGATTTACTTCGATATGTTTAG
 TGCCACTGCATACTTAGTAGCATTATCCATCTGATTGACAGGGGACGCAAAAATATGTATT
 CGCTGATTCTAAACGTCATTTTTGGAACATTGCACATATGGTAAACGCAATTGACCAGAT
 CAATCCTTTTTCTGAAAAGAGGAGCTAACGCTATTGAATTCGACATCGTGTTCATTACAA
 TGGAATAGCGGAACAAACACACCATGGTGTGCCCTGCGACTGCGGACGAATATGCACCAG
 GAAAGAAGACTTCGTCAGTATTTGGATCACATCCGGCTGGTTACTACACCAGATAGCAA
 AGAATTTCTGTGATCAACTGATATTACTTGCACCTGATCTAAAGCTTCAACGAATATCTAC
 AAGCAAAGCCTACGCTGCTGGTGAAGATATAGCAAATAAATCATTGATCATTACTGGCA
 AAAAGGAAATGGTACCGCAAGAGCATAACATATTGCTGAACATCCCAAGTATCCGTCATT
 CGATTTTCATCAACGGTTTTCAACATACAATTATTCGAAGAGAAGGTTATGAAAGATATAA
 TGATAAGTTTGGGATAAACTTCACTGGTAATGATGACCTGGAAGCAACACGTAGGATGCT
 GAAGAGAATGAACATCACCGTAACATTTGGCAAGCTGATGGAATTACCAGCTGTTTCCC
 ACGAGGTACTAGGCGACTGAAAGATGCCATACGTAGAAGAGACACCCCTGGGTACAAAT
 CATTTACAAAAGTATATTCCTGGACTTTGGTCACGTTTTTTGCAATGAGAAGATCAATGAG
 ATTAGGTGTAGATGGTATAATGACCAATCATCCGGAACGTGTAGCATATATTCTTGGAGA
 AAAGTATTTTGTACTACGTTTCAGACTGGCTACTATTGAAGACAACCCGTGGCAGAAAGTA
 TCAGCCCTGATACATCCGTTGGAACAATGAAAAAGGGAGCTGCATTTTTTTGCAAAATG
 TAGAATAATTTTCATATAGTATAAAAATGTTTGTGCATTTGTATATATTTT

>DN125979_c0_g1_i1

TCCAGCAAACCTCTGGTATTTGGTGTGTAATATAGTATATATATAAAGTCATTAACAAAA
 CTTAGCTGATTCCTTCTCTAGTCACCATGTTGTTCTGCACTATACTCTGCATTGCGG
 GGTTGTCTGTGCTAGAGAGGCGAGCTGAGTCAATTAGTGCAGAACGTGATGATGCCAAGC
 GCCCTTTTTATAACATTGCGCACATGGTAAACGGGATAGTTCAAATCAATCAGTTCTTGG
 ATAAAGGAGCTAACGCAATAGAATTCGACTTTGAGTTCAACGACGATGGAATAGCTCAAA
 TAACCCACCACGGATTCCCATGTGACTGCGGAAGAAAATGCAGTTTCCAAGAAAATGTCA
 CCGAATATTTGGATCATATACGGCAGGTCACCTACCCCTGGTATCCCAAATTCGGTGAGC
 AGTTAGTGTTACTCGCGTTGGATTTGAAACTACAAAGGATATTTTGAAGAAAGCACATA
 CTGCTGGGGTGGACGTGGCAAAGAAGCTTCTTGATCATTATTGGCAAAGAGGAAAAAGTA
 AAGCAGAGCCTATTTGCTGCTGAACATCCGTTACTAAAAGACTACGAATTTGTCGAG
 GCTTTAAAGAAACCCTTCGAGACGAAGGTTATGAACAGTACAATGATGTAATAGGCATTA
 ACTTTACGGGTAACGAAGACTTTGGATAAAAACACGCGAAATATAACCAGAAGCTCAATGTCA
 GTAGCCACATCTGGCAATCTGACGGAATTACGAGCTGTTTCCGTAGGAAAATAAGCGCC
 TAAAAGAAGCTTTGAACCTGAGAGATGACCCTAATTACAACATGTTACGAAGGTTTACC
 CGTGGACACTTGTCAAATATACGTGATGAAAGAACGGCTCAGACTGGACGTTGATGGCA
 TACAGACAAACTATCCGGATCGCTTGGATGAAGTCTAAAAGAAAAGGAATTCGCCAGCA
 TGTTCGTTTTGGCAACGTTTTGAAGATGACCCGTGGAAAAAGTTTCATGCGCGCTTAAATAC
 TAGTTGGAGTAACATCACTCACACACATTATTAAGATGAAAATTTCTGTCTGTGGGAGG
 TTCCACATAAATCTCTTCTTCTCAGGATTAATATCCTGCCAAAACCTCCATGGATTT
 TCCTCTGAGACAAATGCAAATGAGAAAAGATTAATAGAGTCATGTTGATTATAAGTAAATA
 TAGATTTGTTTTGCCTATGTGGAAAGCTTTAAAGTATGTAATAATTTCTTAAGAAAATAA
 ATAAAATTTGCAATTAATTATACTTGTAAATAAAGCTTTGTGACTTATTAATTTTCTATGA
 AGTAAATATTATATAAATATATTGTTTACACACATATACAGACAGACATAATGTACCAT
 AATTTATTTTAAATCAAACAGACAACGCATATGAGTTTTTAAATAAATTCCTCTGGTGCTA
 CAAATAAGTCAAACACGTTTTTAAAGTAAAGTTACTAATTATTGAAGCAAATTTTCTTT
 GAAAAGTAAGTTATGTAATCTGTGTGTGTAATAAACACACAATTTTAAAGGGCCCTG
 GTATATTTTGGACGAGCCGTTTCATATCTTTTTTAAAAATTCATCTTAAATCTCATCCGT
 GCTAATATAGTACATGGGGAAAAAAGAAAAGAAAAA

>DN625_c0_g1_i1

TTTTTTTTTTTTCAAGCAGAAGACGGCATAACGAGATACATCGGTGACTGGAGTTCAGACG
 TGTGCTCTCCGATCTGTTTGACTTTGCTCAAATCGAAATGCAATTATTCATTGCTTTGT
 GTTTACTCAGCTCAATTGGGTTACTTGAGGGTGCCAAGCAAGATGGTATCGAACGCACTG
 ATAGCGGCCCTCTATATGGAACATAGCACACATGGTCAATGATAAGGGTCTCATAGATG
 AATATTTGGCTGCTGGAGCAAATAGTGTGAATCTGACGTATCCTTTGATGACAACGGAA
 AACCAATGAAATCCATCATGGCAGGCCTTGTGACTGTGGCAGAAATTGCGATAGGCAGA
 TGAGTTTCACTGACTACTTAGATTATATGCGGCAGCTGACCACTCCTGGCACAGATAATA
 CAATGTTTCGTGAGAATCTAATACTAATTATGCTGGATTTCAAATTGAAGAAACTTTGCG
 CAAATGTAGCGTACACAGCCGGACAGGAGTTGCAAATCAAATGCTGGATACTTATTGGA
 AACGTGGA AAAAAGTGAAGCGAGGGCTTATTTTGTGTTGTCAATACCATCCATCAAAAAG
 CTTGTTTTGCAAAAAGGTTTTACGACAAACTTAACAGCGAAGTTTTGACCAGTATCGGG
 AGAAAGTTGGCGTAGACTTCTCAGGTAACGAGGCCTTTGCAGACACACGTGCAACTTTGT
 CTTGCTGAATATAGTGGATCATATTTGGCAAAGTGATGGCATTACTAATTGCCTCCCTA
 GGAACATGAAGCGTTTGAAGAAGGCGGTAAAGGAAAAGAGATTCCAATGGGTATATGGTTA
 AAGTTTACAACCTGGAGTGTGGACAAGTACACAACAATGAGGAAAAGTCCTACGCGCGGGT
 CTGATGGGATGATAACAAATTTCCCAGGCCGACTTGTTCAGTTCTCAATGAACGTGAAT
 TCTCTGAAAATTTAGGATGGCTACTTACGATGACAATCCATGGGAAAAGTATTAAGCTT
 GAACATGAGTATACTGCATTTAGAGTGACTTGCAACTAACAACAAGTGAATTTTATAA
 TTTTTAGAATTCACCTTAACTATGTTCTTGCTATGGAAGTAAATGCTAAATACATGTAT
 ACTTCTGCCGCAAATAAAGTTTTGATATAACATTTTAAAAAAAAA

>DN101041_c0_g1_i1

TTTTTTTTTTTTTACGAAGTTTTATCAGAATTTTATTTGCAACAGAAAGTACCTCCATAA
 CAAGAACATAAGTACGTAGAATTCTAAGAAATTATAAAATTCACGTGTTATTATTTGCA
 CAAATCCCTTCTAAATGCAGTATACTCTGTGTTAAGCTCAATACCTCTCCCATGGATTG
 TCAGCGTAAGTAGCCAGTCTAAATTTGCCGGAGAATTCACGTTTATTGAGCACTGACACA
 AGTCGGCCGGGAAATTCGTTATCATCCCATCAGCACCGGCGCGTAGGACTTTCCTCATT
 GTTGTGACTTGTCCACACTCCAGTTGTAACCTTTAACCATATACCCATTGGAATCTCTT
 TCCTTTACCGCTTCTTCAAACGCTTTCATGTTCTAGGGAGGCAATTAGTAATGCCATCA
 CTTTGCCAAATATGCTCCACTATATTCAGCGAAGACAAAGTTGCACGTGTGTCTGCAAAG
 GCCTCGTTACCTGAGAAGTCTACGCCAACTTTCTCCCGATACTGGTCAAACCTTCGCTG
 TTAAGTTTGTGCGTAAAACCTTTTGC AAACGAAGCCTTTTATGATGGATGGTATTGACAAC
 ACAAATAAGCCCTCGCTTCACTTTTTCCACGTTTCCAATAAGTATCCAGCATTGATT
 GCAACCTCCTGTCCGGCTGTGTACGCTACATTTGGCGAAAAGTTTCTCAATTTGAAATCC
 AGCATAATTAGTATTAGATTCTCACGAAACATTGTATTATCTGTGCCAGGAGTGGTCAGC
 TGCCGCATATAATCTAAGTAGTCAGTGAACCTCATCTGCCTATCGCAATTTCTGCCACAG
 TCACAAGGCCTGCCATGATGGAATTCATTGGTTTTCCGTTGTCATCAAAGGATACGTCA
 GATTCAACACTATTTGCTCCAGCAGCCAAATATTCATCTATGAGACCCTTATCATTGACC
 ATGTGTGCTATGTTCCATATAGGACGGCCGCTATCAGTGCGTTGATAACCATCTTGCTTG
 GCACCCTCAAGTAACCCAATTGAGCTGAGTAAACACAAAGCAATGAATAATTGCATTTG
 ATTTGAGCAAAGTCAAACGAGAATGATGC

>DN51608_c0_g1_i1

CACCATTTTTACAAAGTGTTAAGAGCAATGTTTATCTTTATTGACATTGATAGAAATGGT
 ATATAAACCCAGGATTACGGACTACCATTGTGACTCAAAGCAAGAAAAGTTCTTTCTAA
 CATGGTGTCTTGTGGCTTTAATGCTTGGTTCTGGGCCGTCGTGCTGGAGGGGTCAGG
 AACAGAGGCTATCAAAAAGCGCGGATGAGCGCCAACCAATATGGAACATAGCAAACATGGT
 AACCGATATCAACCAATTAACGAAACCTCGACCTTGGTGCTAATTCATCGAAACGGA
 TGTAACTTTTGATGATGGGACTATGGCAGGGTACACTTATGAGGGAGTCCCTGCGCCTG
 CTCAGGGTCTGTCACGGAAGAAGATGAATTTGGTGAATTTACACGATTGAGAAGAGT
 CACCACCCCTGGTATCCCATGTATCGCAAAGAGCTTGTCTTAGTTGTGTTCAACTTGAA
 AACAAGTGATTTTTCTCCCTCTACTGCTTACGAGAGTGAATGATGTTCCGGTGGGAACT
 CCTACAAATCTACTGGAACGGAGGGAAGACAGAAAGGAAATGTGATCCTATCAAAGGAAT
 CGTTGTATCTAGAAATGACGGCGGAAAAGCGTACATAATAATATCAATCCCGGACATTGA
 ATATTTTGCAATTCATATCAGGATTTAGGAAAACGCTTGAATTAATGGAATGAAGAAAT
 TCTCGCAAGAGTAGGCTACGAAGTTTCTGGAATTAACCTAGGCTTGATTACAGTATGC
 TTTGAGAAAAGCCGGGTAAGAAGATAGAGCATGTGTGGCAAAGTGATGGTATTACCAA
 CTGCATTTCCACGAGATCTTAATCTCATCGAAGAGCTGTGAAAACAGAGATTTCTCCAGA
 TGGATACATCAACAAAGTGTACCACAGTACGATAGAAGATAAGGCTGCAATCAGTGATGC
 TCTTGACGCTGGTGTGACGGCATAATTTCCCAGAGGCCAGACATGATTTTTGAAGTCTT
 ACTGGATGACCGTTTATAAACCGATTACAGACTAGCCACTTCCGAAGACAACCCGTGGGC

GAATTTTAGATATTGAACAAGCGGACTAATGGTTATTATGATGATATACTTGATGGA AAC
 CTTCTGCGTAAATTTCTTCAGAATCCGCTGTTGAAAAATCATTAGCCTACTCCGTATAGA
 AATAACAATATACGTAGATAAACACTGGCTACAATTTGATGACATGCTTTTAATGTTAGA
 ATAATTTGGTATAAAGCATTGAAGTTTATAAACTTTTATTTTGTATTTGATCGTAAATAA
 AGTTCTATTATTATGTCGCAAAAATGGTTTTGTATTCTGACCATTGAAACGCCGTTTGCA
 AATAAGAAAGTTTTCTTTATGGAAATATAATTTGTTTTGTTTTTAAGTATGTTAAATAG
 TAAGCAGTGGCATTATGTTAGTTCAAAAATGTTCAAAAATTGTTATCAAAAATTAATATTT
 TAAATTCAC

>DN101111_c0_g1_i1

ATATAGTGTATATAGTTAGTTAATATGTTTTAAGCAAATAATTGAGAGACTAAATGGATG
 ATATCTGGAACGGATACTTTTCATTTTCGGCACAAATAAAAAACCCTTAAACGCAATTGCTA
 TTTTTAAGATTACGGGAGCAGAACTTCCTTGAAATAAAAAACATAAAAGAGTGCCTCGC
 AGAATTCCTTTTCTTCTCACAAATTTATTTATAACATGACGACGTTCTGGATACTGATAAA
 TTATTTGTATACTAATCATTATTGACACATAAATTTGTATTTAACTCATGTACATCTCGT
 CCAAGAACAAGAAGTCGGCAGAAATTAATGTAATCCTAACAAACAAATGATAGACTTCTT
 GTAAATCTGCGGAAAATGTGGAATGCATGTGAAAGAATATAAACCAAAAATATCTTCTGC
 TGTGCAAAATTTGACAACCTCAGGATTGTGAAAGACTTTTTAGAGAACAACCTCCAGTACA
 GCTACGCTCCAGATCCTGAACTATCACATTGTAATACCGGTCTGGCGAAGAAAATCATA
 TCACTTCTTCATTGTTACATCGTCTACATCAAGTTCATTTGATACGTCTTCAGTTTCTT
 CAACGTATGTTTTCCAAGGATTGTCATCATACGTTGCCAATTGAATTTGTCTTTGAATT
 CATTTTCTTTGAGGACATCCGCCATATGATTCGAAAAGTTTGTATCACACCGTCAACGC
 TAGCTCTTAAAGCATTCTTATTGTGCGATGCCTGTCTGCAGTTAGAGTGTAAACTTTAT
 TCATATAACCTTGAGAATCTCTTTCTTGCACAGCATCACTTAGTTGCTATTTTCTCGTA
 ATAAGCAATTGGTAATGACCGAACTCTGCCAGACATGCAGTCCAGCCGTACCCAGTCTAT
 TTACAATACGGCTGACGGAGAACAATCATCGTCTCAGCAAAGTTTATTCTACTTTCT
 CGCCATAGGCTGCGAAGCCTCCGTTCTGTAGGGTGAGGTTAAAAGCTCGGAACAACGCAG
 CATTTTTTGTACTTGGTACGGATAGCAGTATATATGCTCTTGCATTACTCTCGGTGTTTT
 CTCTTGGCAGTAATTTCTAAAAGTTTTTCCGCCAAGTTTTTACCAGCTGTATAAGATA
 GACTTGAATTGAAGCCATCAACCTTTAGATCCAATATAATCAGAATCAACTCTTCACGAA
 ACGTTAACTCTCCTGGAGTTGTGAAACGCCGTAATCGTCCAAATAATAACTGAAGTCAC
 GCCATCTGACGCAGTCTCCTGCAAGTCAAGTGGATGGCGTTTTGCTCCGAGATCTAAGAATT
 TCCCAGACGAGTCGAAAATAACCTCAAGTTGGATGGCGTTTTGCTCCGAGATCTAAGAATT
 CGGTGATCCCGATAGGACTGACATCACCTTCTAAAACATGTGCGATTGGATACTCAGCAT
 TCACCGTACGTGCAATGTTCCATATTGGACGTCTCCTGTCTCCCACTTCATCACTGTCTG
 CTCTGTACCCTTCAGTATGAGAATCCAACAGCTAACACCAAGAAGGACGTAAGAACA
 TGCCGAAAATAATTTTCAGCTGCAAACTACAGTATTAGCGTCAAGTTCAAGGGTATATATA
 TGATGATCAACCAACAGATTAACGTTTCGTAAGTACCTAAGTCTATCTAGACAGAAAT
 GTTCGCCACAGCTGCGAAAAGGTGCTTG

>DN101856_c0_g1_i1

AGCACCTTTCGCAGCTGTGGCGAACAATTTCTGTCTAGATAGACTTAGGTACTTACGAAA
 CGTTAATCTGTTGGTTGATCATATATATATATACCCTTGAACCTGACGCTAATACTGTA
 GTTTGCAGCTGAAAATTTTTCGGCATGTTTCTTTACGTCCTTCTTGGTGTAGCTGTT
 GGATTCTCATACTGAAGGGTACAGAAGCAGACAGTATGAAGTGGGAGACAGGAGACGTC
 CAATATGGAAACATTGACGTACGGTGAATGCTGATATCCAATCGACATGTTTTTGAAGA
 GTGATGTCAGTCTATCGGGATCACCGAATTTAGATCTCGGAGCAAACGCCATCCAAC
 TTGAGGTTATTTTCGACTCGTCTGGGAACGCTGAATCTGTTTACCAGAAAACCTCCTTGTG
 ACTGCAGGAGGACCTGCGTCAGATGGCGTGACTTCAGTTATTATTTGGACGATTTACGGC
 GTTTCACAACTCCAGGAGAGTTAACGTTTCGTGAAGAGTTGATTCTGATTATATTGGATC
 TAAAGGTTGATGGCTTCAATTCAAGTCTATCTTATACAGCTGGTAAAAACTTGGCGGAAA
 AACTTTTTAGAAAATTACTGGCAAGAAGAAAACACCGAGAGTAATGCAAGAGCATATATAC
 TGCTATCCGTAACAGTACAAAAAATGCTGCTTGTCCGAGCTTTTAACTCACCCCTAC
 AGAATCCGAGGCTTCGCAGCCTATGGCGAAGTAGGAATAAACTTTGCTGAGGACGATG
 ATTTGTTCTCCGTCAGCCGATTGTAATAGACTGGGTACGGCTGGACTGCATGTCTGGC
 AGAGTTCGGTCATTACCAATTGCTTATTACGAGAAAATAGCGAACTAAGTATGCTGTGTC
 AAGAAAAGAGATTCTCAAGGTTATATGAATAAAGTTTACACTCTAACTGCAGACAGGCATG
 CGACAATAAGAAATGCTTTAAGAGCTAGCGTTGACGGTGTGATGACAACTTTCCGAATC
 ATATGGCGGATGCTCCTCAAAGAAAATGAATTCAAAGACAAATCAAGTTGGCAACGTATG
 ATGACAATCCTTGGAAAACATACGTTGAAGAAAACCTGAAGACGTATCAAAATGAACCTGATG
 TAGACGATGTAACAATGAAGAAGTGATGATTTTTCTTCGCCAGACCGGTATTTACAAT
 GTGATAGTTCAGGATCTGGAGCGTAGCTGTACTGGAAGTTGTTCTCTAAAAGTCTTTTCA
 CAATCCTGAAGTTGTCAAATTTGCACAGCAGAAGATATTTGGTTTATATTCTTTACAT
 GCATCCACATTTTCCCGCAGATTTACAAGAAGTCTATCATTTGTTTGTAGGATTACAT

TAATTTCTGCCGACTTCTTGTCTTGGACGAGATGTACATGAGTTAAATACAAATTTATG
 TGCAATAATGATTAGTATACAAATAATTTATCAGTATCCAGAACGTCGTCATGTTATAA
 ATAAATTGTGAGAAGAAAAGAAAAAAA

9.2 Anexo 2 – Sequências traduzidas dos contigs obtidos através de sequenciamento RNA-seq, através da ferramenta on line Expsy Translate.

>DN101856_c0_g1_i1

MFLYVLLGVSCWILILKGTEADSDEVGDRRRPIWNIARTVNAEYPIDMFLEGDVSPIGITEFLDLGANAIQLEVIFDSS
 GNAESVYQKTPCDCRRTCVWRDFSYLDDLRRFTTPGELTFREELILILDLKVDGFNSSLSTAGKNLAEKLEN
 YWQEENTESNARAYILLSVPSTKNAALFRAFNLTLQNGGFAAYGEKVGINFAEDDDLFVSRIVNRLGTAGLHVWQ
 SSVITNCLLRENSELSDAVQERDSQGYMNKVYTLADRHA TIRNALRASVDGVMTFNPNHMADV LKENEFKDKFK
 LATYDDNPWKTYVEETEDVSNELDVDDVNNEEVI-

>DN101111_c0_g1_i1

MFLYVLLGVSCWILILKGTEADSDEVGDRRRPIWNIARTVNAEYPIDMFLEGDVSPIGITEFLDLGANAIQLEVIFDSS
 GNAESVYQKTPCDCRRTCVWRDFSYLDDLRRFTTPGELTFREELILILDLKVDGFNSSLSTAGKNLAEKLEN
 YWQEENTESNARAYILLSVPSTKNAALFRAFNLTLQNGGFAAYGEKVGINFAEDDDLFVSRIVNRLGTAGLHVWQ
 SSVITNCLLRENSELSDAVQERDSQGYMNKVYTLADRHA TIRNALRASVDGVMTFNPNHMADV LKENEFKDKFK
 LATYDDNPWKTYVEETEDVSNELDVDDVNNEEVI-

>DN51608_c0_g1_i1

MVFLVALMLGSWAVVLEGSGETEAIKSADERQPIWNIANMVT DINQINETLDLGANSIETDVT FDDGT MAGYTYEGV
 LCACSGSCTKKMNI GEFHLHGLRRVTPGDP MYR KELALVVFNLKTSDFSPSTAYESGMMFGGKLLQIYWNGGKTE
 RKCDPIKGIVVSRNDGGKAYIIISIPDIEYFAFISGFRKTL EITGNEEILARVGYEVSGNYLGLIHDALRKAGVKDREH
 VWQSDGITNCIPRDLNLIEEAVKNRDSPDGYINKVYHSTIEDKAAISDALDAGVDGIIISQRPDMIFEVLLDDRFINFR
 LATSEDNPWANFRY-

>DN101041_c0_g1_i1

MQLFIALCLLSSIGLLEGAKQDGIERTDSGRPIWNI AHMVNDKGLIDEYLAAGANSVESDVSFDDNGKPMKFHHGRP
 CDCGRNCDRQMSFTDYLDYMRQLTTPGTDNTMFREN LILIMLDFKLLKLSPNVAYTAGQEVANQMLD TYWKR GK
 SEARAYFVLSIPS IKKASF AKGFYDKLNSEGFDQYREKVGVD FSGNEAFADTRATLSSLNIVDHIWQSDGITNCLPRN
 MKRLKKA VKERDSNGY MVKVYNWSVDKYTTMRKVL RAGADGMITNFPGR LVSVLNEREFSGKFR L ATYADNPW
 ERY-

>DN625_c0_g1_i1

MQLFIALCLLSSIGLLEGAKQDGIERTDSGRPIWNI AHMVNDKGLIDEYLAAGANSVESDVSFDDNGKPMKFHHGRP
 CDCGRNCDRQMSFTDYLDYMRQLTTPGTDNTMFREN LILIMLDFKLLKLSPNVAYTAGQEVANQMLD TYWKR GK
 SEARAYFVLSIPS IKKASF AKGFYDKLNSEGFDQYREKVGVD FSGNEAFADTRATLSSLNIVDHIWQSDGITNCLPRN
 MKRLKKA VKERDSNGY MVKVYNWSVDKYTTMRKVL RAGADGMITNFPGR LVSVLNEREFSGKFR MATYDDNP
 WEKY-

>DN125979_c0_g1_i1

MLFCTILCIAGLSVIEAAESISAERDDAKRPFYNI AHMVNGIVQINQFLDKGANAIEFDFEFNDDGIAQITHHGFPDC
 CGRKC SFQENVT EYLDHIRLVTTPGDPKFR EQLVLLALDLKLRIFSKAHTAGVDVAKKLLDHYWQRGKSKARA
 YLLLNIPLLDYEFVRGFKETLRDEGYEQYNDVIGINFTGNEDL DKTREIYQKLNVS SHIWQSDGITSCFRRKTKRLK
 EALNLRDDPNYNYVTKVYPWTLVKYTSMKERLR LDVDGIQNTY PDR LDEVLKEKEFASMFRLATFEDDPWKKFM
 RA-

>DN5205_c0_g1_i1

MFSATAYLVALSIVLQGTQKYVFADSKRPFWNIAH MVNAIDQINPFLKRGANAIEFDIVFN YNGIAEQTHHGVPDC
 GRICR KEDFVKYLDHIRLVTTPDSKEFRDQLILLALDLKLRISTSKAYAAGEDIANKLIDHYWQKNGGTARAYILL
 NIPSIRHFD FINGFQHTIIRREGYERYNDKFGINFTGNDDLEATRMLKRMNITSNIWQADGITS CFPRGTRRLKDAIR
 RRDTPGYKFIYKVS WTLVTF FAMRRSMRLGVDGIMTNH PERVAYILGEKYFANRFR L ATIEDNPWQKYQP-

>DN28139_c0_g1_i1

MFSATAYLVALSIVLQGTQKYVFADSKRPFWNIAH MVNAIDQINPFLKRGANAIEFDIVFN YNGIAEQTHHGVPDC
 GRICR KEDFVKYLDHIRLVTTPDSKEFRDQLILLALDLKLRISTSKAYAAGEDIANKLIDHYWQKNGGTARAYILL
 NIPSIRHFD FINGFQHTIIRREGYERYNDKFGINFTGNDDLEATRMLKRMNITSNIWQADGITS CFPRGTRRLKDAIR
 RRDTPGYKFIYKVS WTLVTF FAMRRSMRLGVDGIMTNH PERVAYILGEKYFANRFR L ATIEDNPWQKYQP-

>DN82077_c0_g1_i1

MLLSLILVCLIRPSILLDAANGEGFMERAVDGR RPIWNI AHMTNDIEIVR TYLNDGANALEFDIEFNKNGKPVRTFHG
 SPCDCRCKCDRTTTFSTFIEQMRKFTTPGDPEFKKQLVLLMLDLKVTDFDKSLAYTAGQNVAEVLLDNFWQRGKSK
 PQAYLLLSIPSLDYAKLIKGFNDKLDKDEGFVYQEKVGVD FSNDDLDVVRKTLKSIGFESRIWQSDGITNCLPRLTG
 RLKEAIQRDTPGYTYINKVYYWTLDRKSTMRSTLDLGV DGMITNYPDNLNSVLNEEEYVSRFR L ATVEDNPWEIF
 NPSGSRMIEESYDSSLDEEYPDDKLRKRETQDFIPFFLGELFASCAGKALF-

>DN101786_c0_g1_i1

MLLSLILVCLIRPSILLDAANGEGFMERAVDGR RPIWNI AHMTNDIEIVR TYLNDGANALEFDIEFNKNGKPVRTFHG
 SPCDCRCKCDRTTTFSTFIEQMRKFTTPGDPEFKKQLVLLMLDLKVTDFDKSLAYTAGQNVAEVLLDNFWQRGKSK

PQAYLLLSIPSLDYAKLIKGFNDKLDKDEGFVYQEKVGVDFSNDDLDVVRKTLKLSIGFESRIWQSDGITNCLPRLTG
RLKEAIQKRDPGYTYINKVYYWTLDRKSTMRSTLDLGDVDMITNYPDNLNSVLEEEYVSRFRLATVEDNPWEIF
NPSGSRMIEESYDSSLDEEYPDDKLRKRETQDFIPFFLGELFASCAGKALF-

>DN100733_c0_g1_i1

MWLSAVISFIGFAVFLQEANGYVEERADSRKPIWDIAHMVNDLALVDEYLGDGANALEADLAFKSDGTAEKMYH
GTPCDCFRSCSRTEKFDKYMDYIREITTPGSSKFKSGMMLLIIDLKLNIEPNVAYAAGKSTAETLLKKNYWQNGESK
ARAYIVLSLETITRPDFISGFKDAIDASGHTELYEKIGWDFSGNEDLGDIRRAYQKYGIDDHIWQGDGITNCLPRGDY
RLTEAMKKKNDPDYKYTKKVYTWSIDKESSIRNALRLGVDAVMTNYPARVKSVLQESEFSGSHRMATYDDNPWV
K-

>DN100733_c0_g2_i1

MWLSAVISFIGFAVFLQEANGYVEERADSRKPIWDIAHMVNDLALVDEYLGDGANALEADLAFKSDGTAEKMYH
GTPCDCFRSCSRTEKFDKYMDYIREITTPGSSKFKSGMMLLIIDLKLNIEPNVAYAAGKSTAETLLKKNYWQNGESK
ARAYIVLSLETITRPDFISGFKDAIDASGHTELYEKIGWDFSGNEDLGDIRRAYQKYGIDDHIWQGDGITNCLPRGDY
RLTEAMKKKNDPDYKYTKKVYTWSIDKESSIRNALRLGVDAVMTNYPARVKSVLQESEFSGSHRMATYDDNPWV
K-

>DN25784_c0_g1_i1

MLLDISLLLGFCVLLQGAETEVIERADNRRPIWNMGHMVNEIYQIDEFVDLGANSIETDITFDDDAIAEYTYHGVP
CDCKRWCTKWENVDLHGLQRATTPGNSKYHPELVLVVFDLKTGDLSSSTAYKAGNMFAQKLFHIHWNAGNN
GGRAYIVLSIPDIDHYAFISGFREAFKNSDHADLLDKVGYDFSGNDDLSATRNLNKGKGVKDREHVWQSDGITNCI
GRGLGRVRDAVANRDSNGYINKVYVWTIEKYVSVRDAFDAGVDGIMTNEPDVIVDVLNESAYSSKFRMATYEDN
PWETFYK-

>DN545_c0_g1_i1

MIFNIVLTIGFLAVLLEGAETDVSERADNRRPIWNMGHMVNAIYQIDEFVDLGANSLEIDITFDKAKADYTYHKIPC
DCFRTCRKWEYVNDYLKAVRRATTPGYSKYRSRLVLLVFDLKTTKLTNTAYKAGEDFARRLLEHYWNNGNDGG
RAYIVLSIPDVNHYTFINGFGEALRSSGHLELMQKVGDFSGNDDLSTIRNALRKAGVNDKEHIWQSDGITNCLARG
LSRVREAVNDRSSNGYISKVYVWTVDKYKSIRDALDAGVDGVMTPNRPNRVIDVLNENGYSNRFRLATYSDNPFE
TF-

>DN51591_c0_g1_i1

MIFNIVLTIGFLAVLLEGAETDVSERADNRRPIWNMGHMVNAIYQIDEFVDLGANSLEIDITFDKAKADYTYHKIPC
DCFRTCRKWEYVNDYLKAVRRATTPGYSKYRSRLVLLVFDLKTTKLTNTAYKAGEDFARRLLEHYWNNGNDGG
RAYIVLSIPDVNHYTFINGFGEALRSSGHPELMQKVGDFSGNDDLSTIRNALRKAGVNDKEHIWQSDGITNCLARG
LSRVREAVNDRSSNGYISKVYVWTVDKYKSIRDALDAGVDGVMTPNRPNRVIDVLNENGYSNRFRLATYSDNPFE
TF-

>DN100773_c0_g1_i1

MLPYIALVLGCGWGLSESAQTDVGERADKRRPVWIMGHMVNDPPPQIDEFVNLGVNSIETDVSFDDNANPEYTYHG
IPCDCGRSCGHYEFVDFLKNLRIRTPGHSKYHEKLVLVVFDLKTGSLYDNQANDAGKIAKSLQNYWNNGN
GGRAYIVLSIPHLKHYQLITGFQETLKTGHEELLDKVGDFSGNDEISDVENAYKAGVTGHVWQSDGITNCLTR
GLTRIKAAVANRDSNGVINKVYVWTVDKRQSTRDALDAGVDGIMTNPDPITNEIINEAAYKKFRVATYEDNPWE
TFKS-

>DN100812_c0_g1_i1

MYVHLALILGCWTVLQGAETDVGERADNRRPIWNLGHMVNAVVKQIPTFLNDGANAEADITFKGAVPTYSYHGT
PCDFGRDCIRWEYFNVFLQTLRDYTPGNANYREQFILFVLDLKTGSLNNEVRKAGENVAKQLLEHYWNNGNNG
GRAYVVLSPDIAHYEFIRTFKEVLKAEGHENLLDKVGYDLSGPYPLPSLPSLDSVHEAFKAGVDGHVWLSGLTN
WAPLGD MARLKEIVKRRDSENGFISKVYVWSVDKYSTTRTALDVGVDGIMTNFPYVIIDVLNEDGYKDNYRLATY
DDNPWETFKK-

>DN337_c1_g1_i1

MYVHLALILGCWTVLQGAETDVGERADNRRPIWNLGHMVNAVVKQIPTFLDLGANALEADVTFKGSVPTYTYHGT
PCDFGRDCIRWEYFNVFLQTLRDYTPGNANYREQFILFVLDLKTGSLNNEVRKAGENVAKQLLEHYWNNGNNG
GRAYVVLSPDIAHYEFIRTFKEVLKAEGHENLLDKVGYDLSGPYPLPSLPSLDSVHEAFKAGVDGHVWLSGLTN
FAPLLNIDRLKEAIKSRDSANGFINKIYVWSVDKYSTTRTALDVGVDGIMTNYPYVIIDVLNEDGYKDNYRLATY
DNPWETFKKIGRAHV->DN51055_c0_g1_i1

MYVHLALILGCWTVLQGAETDVGERADNRRPIWNLGHMVNAVVKQIPTFLNDGANAEADITFKGAVPTYSYHGT
PCDFGRDCIRWEYFNVFLKTLREYTPGNAKYRDGFILFVLDLKTGSLNDQVRPAGENVAKELLQNYWNNGNDG
GRAYVVLSPDIAHYEFIRTFKEVLKAEGHENLLDKVGYDLSGPYPLPSLPSLDSVHEAFKAGVDGHVWLSGLTN
WAPLGD MARLKEIVKRRDSENGFISKVYVWSVDKYSTTRTALDVGVDGIMTNFPYVIIDVLNEDGYKDNYTLATY
DDNPWETFKK-

>DN405_c0_g1_i1

MLLDISLLLGFCVLLQGAETEVIERADNRRPIWNMGHMVNEIYQIDEFVDLGANSIETDITFDDDAIAEYTYHGVP
CDCKRWCTKWENVDLHGLQRATTPGNSKYHPELVLVVFDLKTGDLSSSTAYKAGNMFAQKLFHIHWNAGNN
GGRAYIVLSIPDIDHYAFISGFREAFKNSDHADLLDKVGYDFSGNDDLSATRNLNKGKGVKDREHVWQSDGITNCI
GRGLGRVRDAVANRDSNGYINKVYVWTIEKYVSVRDAFDAGVDGIMTNEPDVIVDVLNESAYSSKFRMATYEDN
PWETFYK-

9.3 Anexo 3 - Alinhamento entre proteínas do transcriptoma de *L. laeta* peruana e PLDs, de diversas espécies de aranhas do gênero *Loxosceles* e *Sicarius*, do banco de dados do NCBI.

gij292495520|LOXLA/1-276
 DN337_c1_g1_i1/1-317
 gij550540883|LOXLA/1-311
 gij550540884|LOXLA/1-273
 DNS1055_c0_g1_i1/1-311
 gij74814733|LOXLA/1-311
 DN100812_c0_g1_i1/1-311
 gij292495396|LOXRE/1-305
 gij5792507|LOXRE/1-305
 gij41019463|LOXIN/1-315
 gij292495402|LOXIN/1-306
 DN100773_c0_g1_i1/1-305
 gij550540882|LOXGA/1-280
 DNS45_c0_g1_i1/1-306
 DN51591_c0_g1_i1/1-306
 gij209573787|LOXIN/1-307
 DN25784_c0_g1_i1/1-309
 DN405_c0_g1_i1/1-309
 gij292495526|LOXLA/1-276
 gij550540885|LOXIN/1-299
 gij292495518|LOXRU/1-275
 gij292495515|LOXAM/1-275
 DN101856_c0_g1_i1/1-341
 gij1049331759|LOXSI/1-338
 DN101041_c0_g1_i1/1-305
 DN625_c0_g1_i1/1-305
 gij768803645|LOXAR/1-289
 gij122108190|LOXIN/1-305
 DN100733_c0_g2_i1/1-304
 gij49458048|LOXBO/1-276
 gij118573878|LOXIN/1-304
 gij121987412|LOXIN/1-304
 DN82077_c0_g1_i1/1-358
 DN101786_c0_g1_i1/1-358
 DNS205_c0_g1_i1/1-305
 DN28139_c0_g1_i1/1-305
 DN125979_c0_g1_i1/1-308
 gij224472243|SICDA/1-274
 gij292630585|SICCD/1-274
 gij292630589|SICCD/1-274
 gij292630598|SICAL/1-274
 gij292630599|SICDA/1-274

gij292495514|LOXLA/1-276
 gij292495521|LOXLA/1-276
 gij292495520|LOXLA/1-276
 DN337_c1_g1_i1/1-317
 gij550540883|LOXLA/1-311
 gij550540884|LOXLA/1-273
 DNS1055_c0_g1_i1/1-311
 gij74814733|LOXLA/1-311
 DN100812_c0_g1_i1/1-311
 gij292495396|LOXRE/1-305
 gij5792507|LOXRE/1-305
 gij41019463|LOXIN/1-315
 gij292495402|LOXIN/1-306
 DN100773_c0_g1_i1/1-305
 gij550540882|LOXGA/1-280
 DNS45_c0_g1_i1/1-306
 DN51591_c0_g1_i1/1-306
 gij209573787|LOXIN/1-307
 DN25784_c0_g1_i1/1-309
 DN405_c0_g1_i1/1-309
 gij292495526|LOXLA/1-276
 gij550540885|LOXIN/1-299
 gij292495518|LOXRU/1-275
 gij292495515|LOXAM/1-275
 DN101856_c0_g1_i1/1-341
 gij1049331759|LOXSI/1-338
 DN101041_c0_g1_i1/1-305
 DN625_c0_g1_i1/1-305
 gij768803645|LOXAR/1-289
 gij122108190|LOXIN/1-305
 DN100733_c0_g2_i1/1-304
 gij49458048|LOXBO/1-276
 gij118573878|LOXIN/1-304
 gij121987412|LOXIN/1-304
 DN82077_c0_g1_i1/1-358
 DN101786_c0_g1_i1/1-358
 DNS205_c0_g1_i1/1-305
 DN28139_c0_g1_i1/1-305
 DN125979_c0_g1_i1/1-308
 gij224472243|SICDA/1-274
 gij292630585|SICCD/1-274
 gij292630589|SICCD/1-274
 gij292630598|SICAL/1-274
 gij292630599|SICDA/1-274

gij292495514|LOXLA/1-276
 gij292495521|LOXLA/1-276
 gij292495520|LOXLA/1-276
 DN337_c1_g1_i1/1-317
 gij550540883|LOXLA/1-311
 gij550540884|LOXLA/1-273

90 GFDA SRMYK S G K A F E L I Q F S Y W K G S -----DDAGRAY I V L S L P K L D H Y E F I R F K E V H - F D T S 146
 122 S L N N D E V R K A C E N V A K Q L L E H Y W N N G -----N N G G R A Y V V L S L P D I A H Y E F I R F K E V - L K R E 178
 122 S L S N D Q V R P A C E N V A K E L L Q N Y W N N G -----N N G G R A Y V V L S L P D I G H Y E F V R G F K E V - L K K E 178
 84 S L N N H E V R K A C E N V A K G L L E N Y W N N G -----N N G G R A Y V V L S L P D I A H Y E F I R T F K E V - L K T K 140
 122 S L S N D Q V R P A C E N V A K E L L Q N Y W N N G -----N D G G R A Y V V L S L P D I A H Y E F I R T F K E V - L K A E 178
 122 S L N N N E V R K A C E N V A K G L L Q N Y W N N G -----N N G G R A Y V V L S L P D I A H Y E F I R T F K E V - L K A E 178
 122 S L N N N E V R K A C E N I A K G L L K N Y W N N G -----N N G G R A Y V V L S L P D I A H Y E F I R R F K E V - L K A E 178
 123 D F S P S T A Y E S C M M F G G K L L Q I Y W N G G K T E R K C D P I K G I V V S R N D G G K A Y I I I S I P D I E Y F A F I S G F R K T - L E I T 195
 122 S L Y D N Q A Y D A C K L A K N L L K H Y W N N G -----N N G G R A Y I V L S I P D L N H Y K L I T G F K E T - L K S E 178
 122 S L Y D N Q A Y D A C K L A K N L L K H Y W N N G -----N N G G R A Y I V L S I P D L N H Y K L I T G F K E T - L K S E 178
 132 S L Y D N Q A N D A C K L A K N L L Q H Y W N N G -----N N G G R A Y I V L S I P D L N H Y P L I K G F K Q D - L T K D 188
 123 S L Y D N Q A N D A C K L A K N L L Q H Y W N N G -----N N G G R A Y I V L S I P D L N H Y P L I K G F K Q D - L T K D 179
 123 S L Y D N Q A N D A C K I A K S L L Q N Y W N N G -----N N G G R A Y I V L S I P H L K H Y Q L I T G F Q E T - L K T E 179
 97 S L Y D N Q A Y D A C T L A K N L L Q H Y W N N G -----N N G G R A Y I I S I P N L N H Y K L I T G F K E T - L K N E 153
 123 K L T T N T A Y K A C E D F A R R L L E H Y W N N G -----N D G G R A Y I V L S I P D V N H Y T F I N G F G E A - L R S S 179
 123 K L T T N T A Y K A C E D F A R R L L E H Y W N N G -----N D G G R A Y I V L S I P D V N H Y T F I N G F G E A - L R S S 179
 123 Y L T A S T A Y D A C K D F A K R L L Q H Y W N G -----S N G G R A Y I I S I P D L A H Y K F I N G F K E Q - L K T Q 179
 123 D L S S T A Y K A C N M F A Q K L F I H Y W N A G -----N N G G R A Y I V L S I P D I D H Y A F I S G F R E A - F K N S 179
 123 D L S S T A Y K A C N M F A Q K L F I H Y W N A G -----N N G G R A Y I V L S I P D I D H Y A F I S G F R E A - F K N S 179
 90 D L S S T A Y K A C N M F A Q K L F I H Y W N A G -----N N G G R A Y I V L S I P D I D H Y A F I S G F R E A - F K N S 146
 115 D L S S T A Y K G K L F A Q K L L D R Y W N G -----N N G G R A Y I I S I P D I D H Y A F I S G F R E A - L K N A 171
 90 Q L S S S N A Y K G K L F A E K L L R Y W N G -----N N G G R A Y I I S I P K I D H Y A F I S G F R A - L K E S 146
 90 D L S S T A Y K G K L F A E K L L R Y W N G -----S N G G R A Y I I S I P S I D H Y A F I S G F R A - L K K S 146
 136 G F N S L S Y T A C K N L A E K L L E N Y W Q E E T -----E S N A R A Y I L L S V P S T K N A A L F R A F N I L T - L Q N G 194
 136 D F N S L S F K A G K D L A Q K L L S Y W Q K E S -----S G A R A Y I L L S V P S T K N A A L I R A F N I L T - L Q K G 192
 125 K L S P N V A Y T A G E V A N Q M L D T Y W K R G K -----S E A R A Y F V L S I P S I K K A S F A K G F Y D K - L N S E 181
 125 K L S P N V A Y T A G E V A N Q M L D Y W K R G K -----S E A R A Y F V L S I P S I K K A S F A K G F Y D K - L N S E 181
 107 S V A A N L A Y S S G E V A L Q M L N T Y W K R G E -----S G A R A Y I V L S I P T I K R V T F V R G F Y D K - L H S E 163
 123 K L S E Q A Y S A G E V A S Q M L D Y W K R G E -----S G A R A Y I V L S I P T I T R V T F V R G F Y D K - L H S E 179
 123 G I E P N V A Y A A G K S T A E L L K N Y W Q N G E -----S K A R A Y I V L S I L E T I T R P D F I S G F R O A - I D A S 179
 95 G I E P N V A Y A A G K S V A E K L L S Y W Q N G E -----S G A R A Y I V L S I L E T I T R P F I N G F R O A - I K A S 151
 123 G I E P N V A Y A A G K S V A E K L L S G Y W Q N G K -----S G A R A Y I V L S I L E T I T R P F I S G F R O A - I K A S 179
 123 G I E P N V A Y A A G K S V A E K L L S G Y W Q N G K -----S G A R A Y I V L S I L E T I T R P D F I S G F R O A - I K A S 179
 126 F D K S L A Y T A G Q N V A E V L L D N F W Q R G K -----S K P Q A Y L L S I P S L D Y A K L I K G F N D K - L K D E 182
 126 F D K S L A Y T A G Q N V A E V L L D N F W Q R G K -----S K P Q A Y L L S I P S L D Y A K L I K G F N D K - L K D E 182
 120 R I S T S K A Y A A G E D I A N K L I D H Y W Q K G N -----G T A R A Y I L L N I P S I R H F D F I N G F Q H T I I R R E 177
 120 R I S T S K A Y A A G E D I A N K L I D H Y W Q K G N -----G T A R A Y I L L N I P S I R H F D F I N G F Q H T I I R R E 177
 123 R I F S K A H T A G V D A K K L L D H Y W Q R G K -----S K A R A Y L L L N I P S L K D Y E F V R G F K E T - L R D E 179
 90 R I S S A K A Y R A G E D V A K K L L D H Y W Q R G N -----S R A R A Y I L L N I P S V E D Y E F I R A F K D T - L K N E 146
 90 R I S S A K A Y R A G E D V A K K L L D H Y W Q R G N -----S R A R A Y I L L N I P S V E D Y E F I R A F K D T - L K N E 146
 90 R I S S A K A Y R A G E D V A K K L L D H Y W Q R G N -----S R A R A Y I L L N I P S V E D Y E F I R A F K D T - L R N E 146
 90 R I P V E K A Y A A G V D V A T K L L D H Y W Q R G K -----S K A R A Y I L L N I P L V Q D Y E F I R A F K D T - L K N E 146
 90 R I P V E K A Y A A G V D V A T K L L D H Y W Q R G K -----S K A R A Y I L L N I P L V Q D Y E F I R A F K D T - L K N E 146

230 240 250 260 270 280 290

147 A F K D L L L E K V G W D F S G N D D -----L D A T R T A Y Q N A G I E A L N H I W Q S D G T N C I P - R G L G V N K A V S N R D S S D - A 213
 147 T F K N L L E E R V G Y D F S G N E D -----M G L T R V V L D K A G V N D R E H V W Q D G I T N C I L - R S L D R V K A A V A I R D S A T - G 213
 147 T F K N L L E E R V G Y D F S G N E D -----M G L T R V V L N K A G V N D R E H V W Q D G I T N C I L - R S L D R V K A A V A I R D S A T - G 213
 179 G H E N L - L E K V G Y D F S G P Y P S L P T L D Q T H E A F K K A G V E - G H V W Q S D G L T N F A P L N I D R L K E A I K S R D S A N - G 248
 179 G H E D L - L E K V G Y D F S G P Y L P S L P T L D A T H E A Y K K A G V D - G H I W L S D G L T N F S P L G D M A R L K E A I K S R D S A N - G 248
 141 G H E N L - L D K V G Y D L S G P Y P S L P S L D S V H E A F K K A G V D - G H V W L S D G L T N W A L P L G D M A R L K E I V K R R D S E N - G 210
 179 G H E N L - L D K V G Y D L S G P Y L P S L P S L D S V H E A F K K A G V D - G H V W L S D G L T N W A L P L G D M A R L K E I V K R R D S E N - G 248
 179 G H E N L - L D K V G Y D L S G P Y L P S L P S L D S V H E A F K K A G V D - G H V W L S D G L T N W A L P L G D M A R L K E I V K R R D S E N - G 248
 179 G H E N L - L D K V G Y D L S G P Y L P S L P S L D S V H E A F K K A G V D - G H V W L S D G L T N W A L P L G D M A R L K E I V K R R D S E N - G 248
 196 G N E E I - L A R V G Y E V S G N D Y -----L C L I H D A L R K A G V K D R E H V W Q S D G T N C I P - R D L N L I E E A V K N R D S P D - G 261
 179 G H P E L - M D K V G H D F S G N D A -----I C D V G N A Y K K A G V T - G H V W Q S D G T N C L L - R G L S R V R E A V K N R D S S N - G 242
 179 G H P E L - M D K V G H D F S G N D A -----I C D V G N A Y K K A G V T - G H V W Q S D G T N C L L - R G L S R V R E A V K N R D S S N - G 242
 189 G H P E L - M E K V G H D F S G N D D -----I C D V G K A Y K K A G I T - G H I W Q S D G T N C L L - R G L S R V R A A A V A N R D S A N - G 252
 180 G H P E L - M D K V G H D F S G N D D -----I C D V G K A Y K K A G I T - G H I W Q S D G T N C L L - R G L S R V R A A A V A N R D S A N - G 243
 180 G H E E L - L D K V G Y D F S G N D E -----I S D V E N A Y K K A G V T - G H V W Q S D G T N C L L - R G L T I I K A A A V A N R D S N - G 242
 154 G H E E L - L E K V G T D F S G N D D -----I S D V Q T Y N K A G V T - G H V W Q S D G T N C L L - R G L T I V K A A A V A N R D S G S - G 217
 180 G H E L L - M E K V G Y D F S G N D D -----L S T I R N A L R K A G V N D K E H I W Q S D G T N C L A - R G L S R V R E A V K N R D S S N - G 245
 180 G H P E L - M Q K V G Y D F S G N D D -----L S T I R N A L R K A G V N D K E H I W Q S D G T N C L A - R G L S R V R E A V K N R D S S N - G 245
 180 G H E D L - L A K V G Y D F W G N E D -----L S S T R A A F Q K A G V Q D K E H I W Q S D G T N C W L - R T L K R V R E A V N R D S S N - G 245
 180 D H A D L - L D K V G Y D F S G N D D -----L S A T R N A L N K G V K D R E H V W Q S D G T N C I G - R G L G R V R D A V A N R D S S N - G 245
 180 D H A D L - L D K V G Y D F S G N D D -----L S A T R N A L N K G V K D R E H V W Q S D G T N C I G - R G L G R V R D A V A N R D S S N - G 245
 147 D H A D L - L D K V G Y D F S G N D D -----L S A T R N A L N K G V K D R E H V W Q S D G T N C I G - R G L G R V R D A V A N R D S S N - G 212
 172 N H E E L - L D K V G Y D F S G N D D -----L S S T R A L N K A G V K D R E H V W Q S D G T N C I A - R G L D R V R E A V R N D S S N - G 237
 147 K H E E L - L A K V G Y D F S G N D D -----L D S I R S A L N K A G V K D R E H V W Q S D G T N C I A - R N L N R V R E A V Q N R D S S N - G 212
 147 G H E D L - L A K V G Y D F S G N D D -----L N S I R S A L H K A G V K D R E H V W Q S D G T N C L L A - R S L D V L N E A V K N R D S S N - G 212
 195 G F A A Y - G E K V G I N F A E D D D -----L F S V S R I V N R L G T A - G L H V W Q S S V T N C L L - R E N S E L S D A V Q E R D S Q G Y - 259
 193 E F A F F - G E K V G I D I S G D D D -----L F S V R R I V E R L G M T - G W H V W Q S D V T N C L L - R E K S O L R E A V E E R D S Q G Y - 257
 182 G F D Q Y - R E K V G V D F S G N E A -----F A D T R A T L S S L N I V - D H I W Q S D G T N C L P - R N M K L K K A V K E R D S N G Y - 245
 182 G F D Q Y - R E K V G V D F S G N E A -----F A D T R A T L S S L N I V - D H I W Q S D G T N C L P - R N M K L K K A V K E R D S N G Y - 245
 164 G F D Q Y - R E K V G V D F S G N E D -----L D E T G R I L S S Q N I L - D H I W Q S D G T N C I F - R V M T R L K K A I I N K R D S N G Y - 227
 180 G F D Q Y - R E K V G V D F S G N E D -----L E D T G K I L K S R D I L - D H I W Q S D G T N C L P - R I M K R L K A A I I R K R D S N G Y - 243
 180 G H T E L - Y E K I C W D F S G N E D -----L G D I R R A Y Q K Y G I D - D H I W Q D G G T N C L P - R G D Y L T E A M K K N D P D Y K 244
 152 G H E E L - F E K I C W D F S G N E D -----L G D I R R V Y Q K Y G I D - E H I W Q D G G T N C L P - R G D Y L T E A M K K N D P D Y K 216
 180 G H E E L - F E K I C W D F S G N E D -----L G E I R R V Y Q K Y G I E - D H I W Q D G G T N C L P - R G D Y L T E A M K K N D P N Y K 244
 180 G H E E L - F E K I C W D F S G N E D -----L G E I R R V Y Q K Y G I E - D H I W Q D G G T N C L P - R G D Y L T E A M K K N D P N Y K 244
 183 G F E V Y - Q E K V G V D F S N N D D -----L D D V R K T L K S I G F E - S R I W Q S D G T N C L P - R L T G R L K E A I Q K R D T P G Y T 247
 183 G F E V Y - Q E K V G V D F S N N D D -----L D D V R K T L K S I G F E - S R I W Q S D G T N C L P - R L T G R L K E A I Q K R D T P G Y T 247
 178 G Y E R Y - N D K F G I N F T G N D D -----L E A T R R M L K R M N I T - S N I W Q A D G T S C F P - R G T R L K D A I I R R R D T P G Y K 242
 178 G Y E R Y - N D K F G I N F T G N D D -----L E A T R R M L K R M N I T - S N I W Q A D G T S C F P - R G T R L K D A I I R R R D T P G Y K 242
 180 G Y E S Y - N D V I G I N F T G N E D -----L D K I R D V L E I L G I H - K Q V W Q A D G T S C F A - R G T E R L K E A L E K R D T P G Y N 211
 147 G Y E S Y - N D K V G I N F T G N E D -----L D K I R D V L E I L G I H - K Q V W Q A D G T S C F A - R G T E R L K E A L E K R D T P G Y N 211
 147 G Y E S Y - N D K V G I N F T G N E D -----L D K I R D V L E I L G I H - K Q V W Q A D G T S C F A - R G T E R L K E A L E K R D T P G Y N 211
 147 G Y E Q Y - N D K V G V N F T G N E D -----L D E I R K V L E K V G V D - K H V W Q A D G T S C F T - R G T D R L T E A L K R R D T P G Y N 211
 147 G Y E Q Y - N D K V G V N F T G N E D -----L D E I R K V L E K V G V D - K H V W Q A D G T S C F A - R G T D R L T E A L K R R D T P G Y N 211

300 310 320 330 340 350 360

214 F I N K V Y V W T E K Y S S V K D A L S A D V G C I M T N H P N V I N G V L K E D E F K D R F K L A T Y E D N P W T F K R ----- 276
 214 Y I N K V Y F W T I Q A Y S S V R D A L N A E V D G C I M T N E P D V I A N I L K E D A F K D R F R L A T Y R D N P W E T F K R ----- 276
 214 Y I N K V Y F W T I Q A Y S S V R D A L N A E V D G C I M T N E P D V I A N V L K E D A F K D R F R L A T Y R D N P W E T F K R ----- 276
 249 F I N K I Y Y V S V D K Y S T T R T A L D V G V D C I M T N Y P V I I D V L N E D G Y K D K Y R L A T Y D D N P W E T F K I G R A H V ----- 317
 249 F I N K I Y Y V S V D K Y S T T R T A L D V G V D C I M T N Y P V I I D V L N E D G Y K D K Y R L A T Y D D N P W E T F K I G R A H V ----- 311
 211 F I S K V Y V S V D K Y S T T R T A L D V G V D C I M T N Y P V I I D V L N E A K Y K Y R I A T Y R D N P W E T F K N ----- 273

g|122108190|LOXIN/1-305
 DN100733_c0_g1_1/1-311
 DN100812_c0_g1_1/1-311
 g|174814733|LOXLA/1-311
 DN51608_c0_g1_1/1-324
 g|12924955396|LOXRE/1-305
 g|157792507|LOXRE/1-305
 g|141019463|LOXIN/1-315
 g|1292495402|LOXIN/1-306
 DN100773_c0_g1_1/1-305
 g|1550540882|LOXGA/1-280
 DN545_c0_g1_1/1-306
 DN51591_c0_g1_1/1-306
 g|1209573787|LOXIN/1-307
 DN25784_c0_g1_1/1-309
 DN405_c0_g1_1/1-309
 g|1292495526|LOXLA/1-276
 g|1550540885|LOXIN/1-299
 g|1292495518|LOXRU/1-275
 g|1292495515|LOXAM/1-275
 DN101856_c0_g1_1/1-341
 g|11049331759|LOXS/1-338
 DN101041_c0_g1_1/1-305
 DN625_c0_g1_1/1-305
 g|1768803645|LOXAR/1-289
 g|1122108190|LOXIN/1-305
 DN100733_c0_g2_1/1-304
 g|149458048|LOXBO/1-276
 g|118573878|LOXIN/1-304
 g|1121987412|LOXIN/1-304
 DN82077_c0_g1_1/1-358
 DN101786_c0_g1_1/1-358
 DN5205_c0_g1_1/1-305
 DN28139_c0_g1_1/1-305
 DN125979_c0_g1_1/1-308
 g|1224472243|SCDA/1-274
 g|1292630585|SCCD/1-274
 g|1292630589|SCCD/1-274
 g|1292630598|SCAL/1-274
 g|1292630599|SCDA/1-274

249 F I S K V Y Y S V D K Y S T T R T A L D V G D G I M T N F P Y V I I D V L N E D G Y K D N Y T L A T Y D D N P W E T F K K ----- 311
 249 F I S K V Y Y S V D K Y S T T R T A L D V G D G I M T N F P Y V I I D V L N E D G Y K D N Y T L A T Y D D N P W E T F K K ----- 311
 249 F I S K V Y Y S V D K Y S T T R T A L D V G D G I M T N F P Y V I I D V L N E N G Y K D K Y R L A T Y D D N P W E T F K K ----- 311
 262 Y I N K V Y Y S T I E D K A A I S D A L D A G V D G I I S Q R P D M I F E V L L D D R F I N R F R L A T S E D N P W A N F R Y ----- 324
 243 F I N K V Y Y S T V D K R A T T R E A L D A C V D G V M T N Y P D V I T D V L N E S A Y K A K F R I A T Y D D N P W E T F K N ----- 305
 243 F I N K V Y Y S T V D K R S T T R D A L D A C V D G I M T N Y P D V I T D V L N E S A Y K A K F R V A T Y D D N P W T F K K ----- 305
 253 F I N K V Y Y S T V D K R S T T R D A L D A C V D G I M T N Y P D V I T D V L N E S A Y K A K F R V A T Y D D N P W T F K K ----- 315
 244 F I N K V Y Y S T V D K R S T T R D A L D A C V D G I M T N Y P D V I T D V L N E S A Y K A K F R V A T Y D E N P W T F K K ----- 306
 243 V I N K V Y Y S T V D K R Q S T R D A L D A C V D G I M T N Y P D I T N E I L N E S A Y K A K F R V A T Y E D N P W E T F K S ----- 305
 218 I N K V Y Y S T V D K R Q S T R D T L D A N V D G I M T N Y P D I T V E I L N E S A Y K A K F R I A T Y E D N P W E T F K G ----- 280
 246 Y I S K V Y Y S T V D K Y K S I R D A L D A C V D G V M T N Y P N R V I D V L N E N G Y S N R F R L A T Y S D N P E E T F ----- 306
 246 Y I S K V Y Y S T V D K Y K S I R D A L D A C V D G V M T N Y P N R V I D V L N E N G Y S N R F R L A T Y S D N P E E T F ----- 306
 246 Y I N K V Y Y S T V D K Y S V R D A I N A G A D G I M T N Y P N V I V D V L K E N D F K G K F R M A T Y N D N P W E T F K ----- 307
 246 Y I N K V Y Y S T I E K Y V S V R D A F D A C V D G I M T N E P D V I V D V L N E S A Y S S K F R M A T Y E D N P W E T F K Y K ----- 309
 246 Y I N K V Y Y S T I E K Y V S V R D A F D A C V D G I M T N E P D V I V D V L N E S A Y S S K F R M A T Y E D N P W E T F K Y K ----- 309
 246 Y I N K V Y Y S T I E K Y V S V R D A F D A C V D G I M T N E P D V I V D V L N E S A Y S S K F R M A T Y E D N P W E T F K Y K ----- 309
 238 Y I N K V Y Y S T I E K Y V S V R D A L D A C V D G I M T N E P D V I V N V L N E G N Y R G R F R L A N Y D D N P W T F K ----- 299
 213 Y I N K V Y Y S T I E K Y V S V R D A L N A E V D G I M T N Y P N V I V N V L N E D N F R N R F R M A T F E D N P W E H F T R ----- 275
 213 Y I S K M Y Y S T I E K Y A T V R D A L N A E V D G I M T N Y P D V I V N V L N E D S F K N R F R M A T F D D N P W E L F K R ----- 275
 260 -M N K Y Y T S T A D R H A T I R N A L R A C V D G V M T N F P N H M A D V L K E N E F K D K F K L A T Y D D S P W E T Y M E D T E Y I S N T L N V ----- 332
 258 -M N K Y Y T S T A D R H A T I R N A L R A C V D G V M T N F P N H L T D V L K E T E F K D K F K L A T Y D D S P W E T Y M E D T E Y I S N T L N V ----- 330
 246 -M V K V Y Y S V D K Y T T M R K V L R A C A D G M I T N F P G R L V S V L N E R E F S G K F R L A T Y A D N P W E R Y ----- 305
 246 -M V K V Y Y S V D K Y T T M R K V L R A C A D G M I T N F P G R L V S V L N E R E F S G K F R M A T Y D D N P W E K Y ----- 305
 228 -M V K V Y Y S V D K Y T T M R K T L R A C A D G M I T N F P D R L V S V L N E R E F S G K F R L A T Y D D N P W E R Y K A ----- 289
 244 -M V K V Y Y S V D K Y T T M R K A L R A C A D G M I T N F P K R L V S V L N E R E F S G K F R L A T Y D D N P W E R Y T G ----- 305
 245 Y T K K Y Y T S I D K E S S I R N A L R L C V D A V M T N Y P A R V K S I L R E S E F S G S H R M A T Y D D N P W K ----- 304
 217 Y T E K Y Y T S I D K E S S I R N A L R L C V D A V M T N Y P A R V K S I L R E S E F S S T H R M A T Y D D N P W K ----- 276
 245 Y T L K Y Y T S I D K E S S I R N A L R L C V D A V M T N Y P A R V K S I L R E S E F S G T H R M A T Y D D N P W K ----- 304
 245 Y T L K Y Y T S I D K E S S I R N A L R L C V D A V M T N Y P A R V K S I L R E S E F S G T H R M A T Y D D N P W K ----- 304
 248 Y I N K V Y Y S T L D R K S T M R S T L D L C V D G M I T N Y P D N L N S V L N E E E Y V S R F R L A T Y E D N P W E I F N P S G S R M I E E S Y D ----- 321
 248 Y I N K V Y Y S T L D R K S T M R S T L D L C V D G M I T N Y P D N L N S V L N E E E Y V S R F R L A T Y E D N P W E I F N P S G S R M I E E S Y D ----- 321
 243 F I Y K V Y S T M L V T F F A M R R S M R L C V D G I M T N H P E R V A Y I L C E K Y F A N R F R L A T I E D N P W Q Y Q P ----- 305
 243 F I Y K V Y S T M L V T F F A M R R S M R L C V D G I M T N H P E R V A Y I L C E K Y F A N R F R L A T I E D N P W Q Y Q P ----- 305
 245 Y V T K V Y P M T L V K Y T S M K E R L R D V D G I Q T N Y P D R L D E V L K E K E F A S M F R L A T F E D D P W K Y F M R A ----- 308
 212 Y I N K V Y A M T L V R K S I M R R S L R L C V D G V M S N N P D R V I K V L K E E F A D K F R L A T Y D D N P W E K F R G ----- 274
 212 Y I N K V Y A M T L V R K S I M R R S L R L C V D G V M S N N P D R V I K V L K E E F A D K F R L A T Y D D N P W E K F R G ----- 274
 212 Y I N K V Y A M T L V R K S I M R R S L R L C V D G V M S N N P D R V I K V L K E E F A D K F R L A T Y D D N P W E K F R G ----- 274
 212 Y A Y K V A M T L V K Y S T M R R S L R L C V D G V M S N F P D R V V E V L K E E E F A G K F R M A T Y D D N P W K F T S ----- 274
 212 Y A Y K V A M T L V K Y S T M R R S L R L C V D G V M S N F P D R V V E V L K E E E F A D K F R M A T Y D D N P W K F T G ----- 274

380 390 400

g|1292495514|LOXLA/1-276
 g|1292495521|LOXLA/1-276
 g|1292495520|LOXLA/1-276
 DN337_c1_g1_1/1-317
 g|1550540883|LOXLA/1-311
 g|1550540884|LOXLA/1-273
 DN51055_c0_g1_1/1-311
 DN100812_c0_g1_1/1-311
 g|174814733|LOXLA/1-311
 DN51608_c0_g1_1/1-324
 g|12924955396|LOXRE/1-305
 g|157792507|LOXRE/1-305
 g|141019463|LOXIN/1-315
 g|1292495402|LOXIN/1-306
 DN100773_c0_g1_1/1-305
 g|1550540882|LOXGA/1-280
 DN545_c0_g1_1/1-306
 DN51591_c0_g1_1/1-306
 g|1209573787|LOXIN/1-307
 DN25784_c0_g1_1/1-309
 DN405_c0_g1_1/1-309
 g|1292495526|LOXLA/1-276
 g|1550540885|LOXIN/1-299
 g|1292495518|LOXRU/1-275
 g|1292495515|LOXAM/1-275
 DN101856_c0_g1_1/1-341
 g|11049331759|LOXS/1-338
 DN101041_c0_g1_1/1-305
 DN625_c0_g1_1/1-305
 g|1768803645|LOXAR/1-289
 g|1122108190|LOXIN/1-305
 DN100733_c0_g2_1/1-304
 g|149458048|LOXBO/1-276
 g|118573878|LOXIN/1-304
 g|1121987412|LOXIN/1-304
 DN82077_c0_g1_1/1-358
 DN101786_c0_g1_1/1-358
 DN5205_c0_g1_1/1-305
 DN28139_c0_g1_1/1-305
 DN125979_c0_g1_1/1-308
 g|1224472243|SCDA/1-274
 g|1292630585|SCCD/1-274
 g|1292630589|SCCD/1-274
 g|1292630598|SCAL/1-274
 g|1292630599|SCDA/1-274

333 ---D V N N E E V I ----- 341
 331 ---D D I S I A K K ----- 338

322 S L L D E E Y P D D K L K R E T Q D F I P F F L G E L F A S C A C K A L F ----- 358
 322 S L L D E E Y P D D K L K R E T Q D F I P F F L G E L F A S C A C K A L F ----- 358