Data in Brief 15 (2017) 272–280

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Salivary lipocalin family proteins from *Panstrongylus chinai*, a vector of Chagas disease

Hirotomo Kato [a],*, Ryan C. Jochim [b], Eduardo A. Gomez [c],
Shunsuke Tsunekawa [d], Jesus G. Valenzuela [b],
Yoshihisa Hashiguchi [c]

[a] *Division of Medical Zoology, Department of Infection and Immunity, Jichi Medical University, Shimotsuke, Tochigi, Japan*
[b] *Vector Molecular Biology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, NIH, Rockville, MD, USA*
[c] *Departamento de Parasitologia y Medicina Tropical, Centro de Biomedicina, Facultad de Ciencias Medicas, Universidad Catolica de Santiago de Guayaquil, Guayaquil, Ecuador*
[d] *Laboratory of Parasitology, Department of Disease Control, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo, Hokkaido, Japan*

ARTICLE INFO

Article history:
Received 24 August 2017
Received in revised form
7 September 2017
Accepted 19 September 2017
Available online 22 September 2017

Keywords:
Panstrongylus chinai
Saliva
Lipocalin
Transcriptome

ABSTRACT

The dataset in this report is related to the research article with the title: “Salivary gland transcripts of the kissing bug, *Panstrongylus chinai*, a vector of Chagas disease” (Kato et al., 2017) [1]. Lipocalin family proteins were identified as the dominant component in *P. chinai* saliva, and phylogenetic analysis of the salivary lipocalins resulted in the formation of five major clades. For further characterization, each clade of *P. chinai* lipocalin was s alignment and phylogenetic analyses together with homologous triatomine lipocalins; pallidipin 2, an inhibitor of collagen-induced platelet aggregation identified from saliva of *Triatoma pallidipennis* (clade I), pallidipin-like salivary lipocalin from *Triatoma dimidiata* (clade II), salivary lipocalin from *T. dimidiata* (clade III), triatin-like salivary lipocalin identified in the saliva of *T. dimidiata* (clade IV), and lipocalin-like TiLipo37 from *Triatoma infestans* (clade V).

DOI of original article: http://dx.doi.org/10.1016/j.actatropica.2017.06.022

* Corresponding author.
E-mail address: hirok@jichi.ac.jp (H. Kato).

http://dx.doi.org/10.1016/j.dib.2017.09.039

Specifications Table

Subject area	*Biology*
More specific subject area	*Salivary lipocalins of a hematophagous insect*
Type of data	*Table, figure*
How data was acquired	*Transcriptome, alignment and phylogenetic analyses*
Data format	*Analyzed*
Experimental factors	*A dataset of transcripts from salivary grands of Panstrongylus chinai*
Experimental features	*Alignment and phylogenetic analyses of salivary lipocalins from Panstrongylus chinai*
Data source location	*Ecuador*
Data accessibility	*Accession numbers of the sequence data are available in the reference [1].*

Value of the data

- The data is the second report of the salivary lipocalins in a *Panstrongylus* species.
- The result will provide further information into the salivary biochemical and pharmacological complexity of triatomine bugs and the evolution of salivary components in blood sucking arthropods.
- cDNAs and recombinant proteins prepared from these transcripts will result in the discovery of novel pharmacologically active compounds, as well as the development of biomarkers following exposure to *Panstrongylus chinai*.

1. Data

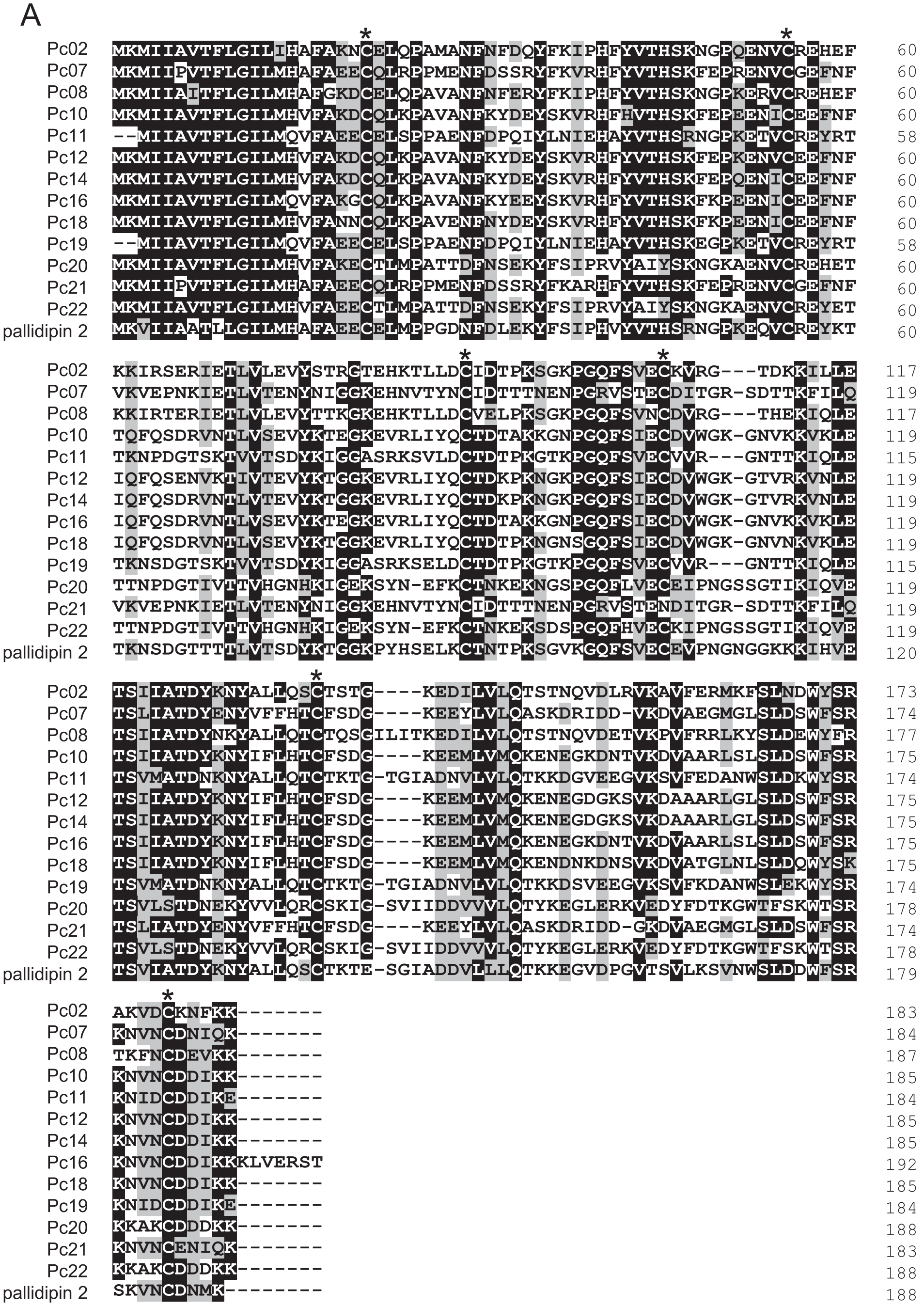
The dominant transcripts of *Panstrongylus chinai* salivary glands were analyzed by sequence analysis of the cDNA library, and 73.7% of transcripts encoding the putative secreted proteins coded for the lipocalin family of proteins [1]. Table 1 shows the grouping of transcripts coding for lipocalin family proteins in *P. chinai* salivary glands obtained by the phylogenetic analysis [1]. Figs. 1–5 represent alignment and phylogenetic analyses of each clade of *P. chinai* salivary lipocalins together with homologous proteins; pallidipin 2, a platelet aggregation inhibitor identified from *Triatoma pallidipennis* saliva (clade I), pallidipin-like salivary lipocalin from *Triatoma dimidiata* saliva with unknown function (clade II), salivary lipocalin from *T. dimidiata* saliva with unknown function (clade III), triatin-like salivary lipocalin identified in the saliva of *T. dimidiata* with unknown function (clade IV), and lipocalin-like TiLipo37 from *Triatoma infestans* saliva with unknown function (clade V), showing their structural similarity and diversity.

2. Experimental design, materials and methods

The sequences of *P. chinai* salivary lipocalins were obtained in the study “Salivary gland transcripts of the kissing bug, *Panstrongylus chinai*, a vector of Chagas disease” [1]. The sequences coding for lipocalin family of proteins by BLASTx analyses were aligned with CLUSTAL W software [2] and examined using Molecular Evolutionary Genetics Analysis (MEGA) version 5.2 [3]. Phylogenetic trees by the neighbor-joining method were constructed with the distance algorithms available in the MEGA package. Bootstrap values were determined on 1000 replicates of the data sets. Accession numbers of the sequence data are available in “Salivary gland transcripts of the kissing bug, *Panstrongylus chinai*, a vector of Chagas disease” [1].

Table 1
Transcripts coding for lipocalin family proteins in *Panstrongylus chinai* salivary glands.

Clade	Similar to	No. of clusters	No. of seq	% seq
Clade I				
	pallidipin 2 (*Triatoma pallidipennis*): AAA30329	16	151	46.9
	Td42, similar to pallidipin 2 (*Triatoma dimidiata*): BAI50842	1	1	0.3
Clade II				
	Td38, similar to pallidipin-like salivary lipocalin (*Triatoma dimidiata*): BAI50839	2	5	1.6
	Td33, similar to pallidipin-like salivary lipocalin (*Triatoma dimidiata*): BAI50837	1	1	0.3
Clade III				
	Td26, similar to salivary lipocalin (*Triatoma dimidiata*): BAI50831	2	11	3.4
	salivary lipocalin (*Triatoma infestans*): ABR27920	1	8	2.4
	Td40, similar to triabin-like lipocalin 4a precursor (*Triatoma dimidiata*): BAI50840	1	3	0.9
	triabin-like lipocalin 4a precursor (*Triatoma infestans*): ABR27959	1	2	0.7
	salivary lipocalin (*Triatoma infestans*): ABR27868	1	1	0.3
Clade IV				
	Td18, similar to triatin-like salivary lipocalin (*Triatoma dimidiata*): BAI50824	4	66	20.5
	Td11, similar to triatin-like salivary lipocalin (*Triatoma dimidiata*): BAI50818	1	43	13.3
	Td45, similar to pallidipin-like salivary lipocalin (*Triatoma dimidiata*): BAI50844	1	8	2.5
Clade V				
	lipocalin-like TiLipo37 (*Triatoma infestans*): AAQ68063	2	10	3.1
	salivary lipocalin (*Triatoma brasiliensis*): ABH09436	1	6	1.9
Others				
	Td24, similar to salivary lipocalin (*Triatoma dimidiata*): BAI50829	2	2	0.7
	triabin-like salivary lipocalin (*Triatoma infestans*): ABR27927	1	1	0.3
	salivary lipocalin (*Triatoma infestans*): ABR27831	1	1	0.3
	salivary lipocalin 1 (*Triatoma brasiliensis*): ABH09421	1	1	0.3
	Td23, similar to salivary lipocalin (*Triatoma dimidiata*): BAI50828	1	1	0.3
Total		41	322	100.0



```
                                 *                                 *
Pc02         MKMIIAVTFLGILIHAFAKNCELQPAMANFNFDQYFKIPHFYVTHSKNGPQENVCREHEF  60
Pc07         MKMIIPVTFLGILMHAFAEECQLRPPMENFDSSRYFKVRHFYVTHSKFEPRENVCGEFNF  60
Pc08         MKMIIAITFLGILMHAFGKDCELQPAVANFNFERYFKIPHFYVTHSKNGPKERVCREHEF  60
Pc10         MKMIIAVTFLGILMHVFAKDCQLKPAVANFKYDEYSKVRHFHVTHSKFEPEENICEEFNF  60
Pc11         --MIIAVTFLGILMQVFAEECELSPPAENFDPQIYLNIEHAYVTHSRNGPKETVCREYRT  58
Pc12         MKMIIAVTFLGILMHVFAKDCQLKPAVANFKYDEYSKVRHFYVTHSKFEPKENVCEEFNF  60
Pc14         MKMIIAVTFLGILMHVFAKDCQLKPAVANFKYDEYSKVRHFYVTHSKFEPQENICEEFNF  60
Pc16         MKMIIAVTFLGILMQVFAKGCQLKPAVANFKYEEYSKVRHFYVTHSKFKPEENICEEFNF  60
Pc18         MKMIIAVTFLGILMHVFANNCQLKPAVENFNYDEYSKVRHFYVTHSKFKPEENICEEFNF  60
Pc19         --MIIAVTFLGILMQVFAEECELSPPAENFDPQIYLNIEHAYVTHSKEGPKETVCREYRT  58
Pc20         MKMIIAVTFLGILMHVFAKECTLMPATTDFNSEKYFSIPRVYAIYSKNGKAENVCREHET  60
Pc21         MKMIIPVTFLGILMHAFAEECQLRPPMENFDSSRYFKARHFYVTHSKFEPRENVCGEFNF  60
Pc22         MKMIIAVTFLGILMHVFAEECTLMPATTDFNSEKYFSIPRVYAIYSKNGKAENVCREYET  60
pallidipin 2 MKVIIAATLLGILMHAFAEECELMPPGDNFDLEKYFSIPHVYVTHSRNGPKEQVCREYKT  60

                                         *               *
Pc02         KKIRSERIETLVLEVYSTRGTEHKTLLDCIDTPKSGKPGQFSVECKVRG---TDKKILLE  117
Pc07         VKVEPNKIETLVTENYNIGGKEHNVTYNCIDTTTNENPGRVSTECDITGR-SDTTKFILQ  119
Pc08         KKIRTERIETLVLEVYTTKGKEHKTLLDCVELPKSGKPGQFSVNCDVRG---THEKIQLE  117
Pc10         TQFQSDRVNTLVSEVYKTEGKEVRLIYQCTDTAKKGNPGQFSIECDVWGK-GNVKKVKLE  119
Pc11         TKNPDGTSKTVVTSDYKIGGASRKSVLDCTDTPKGTKPGQFSVECVVR---GNTTKIQLE  115
Pc12         IQFQSENVKTIVTEVYKTGGKEVRLIYQCTDKPKNGKPGQFSIECDVWGK-GTVRKVNLE  119
Pc14         IQFQSDRVNTLVTEVYKTGGKEVRLIYQCTDKPKNGKPGQFSIECDVWGK-GTVRKVNLE  119
Pc16         IQFQSDRVNTLVSEVYKTEGKEVRLIYQCTDTAKKGNPGQFSIECDVWGK-GNVKKVKLE  119
Pc18         IQFQSDRVNTLVSEVYKTGGKEVRLIYQCTDTPKNGNSGQFSIECDVWGK-GNVNKVKLE  119
Pc19         TKNSDGTSKTVVTSDYKIGGASRKSELDCTDTPKGTKPGQFSVECVVR---GNTTKIQLE  115
Pc20         TTNPDGTIVTTVHGNHKIGEKSYN-EFKCTNKEKNGSPGQFLVECEIPNGSSGTIKIQVE  119
Pc21         VKVEPNKIETLVTENYNIGGKEHNVTYNCIDTTTNENPGRVSTENDITGR-SDTTKFILQ  119
Pc22         TTNPDGTIVTTVHGNHKIGEKSYN-EFKCTNKEKSDSPGQFHVECKIPNGSSGTIKIQVE  119
pallidipin 2 TKNSDGTTTTLVTSDYKTGGKPYHSELKCTNTPKSGVKGQFSVECEVPNGNGGKKKIHVE  120

                             *
Pc02         TSIIATDYKNYALLQSCTSTG----KEDILVLQTSTNQVDLRVKAVFERMKFSLNDWYSR  173
Pc07         TSLIATDYENYVFFHTCFSDG----KEEYLVLQASKDRIDD-VKDVAEGMGLSLDSWFSR  174
Pc08         TSIIATDYNKYALLQTCTQSGILITKEDILVLQTSTNQVDETVKPVFRRLKYSLDEWYFR  177
Pc10         TSIIATDYKNYIFLHTCFSDG----KEEMLVMQKENEGKDNTVKDVAARLSLSLDSWFSR  175
Pc11         TSVMATDNKNYALLQTCTKTG-TGIADNVLVLQTKKDGVEEGVKSVFEDANWSLDKWYSR  174
Pc12         TSIIATDYKNYIFLHTCFSDG----KEEMLVMQKENEGDGKSVKDAAARLGLSLDSWFSR  175
Pc14         TSIIATDYKNYIFLHTCFSDG----KEEMLVMQKENEGDGKSVKDAAARLGLSLDSWFSR  175
Pc16         TSIIATDYKNYIFLHTCFSDG----KEEMLVMQKENEGKDNTVKDVAARLSLSLDSWFSR  175
Pc18         TSIIATDYKNYIFLHTCFSDG----KEEMLVMQKENDNKDNSVKDVATGLNLSLDQWYSK  175
Pc19         TSVMATDNKNYALLQTCTKTG-TGIADNVLVLQTKKDSVEEGVKSVFKDANWSLEKWYSR  174
Pc20         TSVLSTDNEKYVVLQRCSKIG-SVIIDDVVVLQTYKEGLERKVEDYFDTKGWTFSKWTSR  178
Pc21         TSLIATDYENYVFFHTCFSDG----KEEYLVLQASKDRIDD-GKDVAEGMGLSLDSWFSR  174
Pc22         TSVLSTDNEKYVVLQRCSKIG-SVIIDDVVVLQTYKEGLERKVEDYFDTKGWTFSKWTSR  178
pallidipin 2 TSVIATDYKNYALLQSCTKTE-SGIADDVLLLQTKKEGVDPGVTSVLKSVNWSLDDWFSR  179

                 *
Pc02         AKVDCKNFKK-------                                            183
Pc07         KNVNCDNIQK-------                                            184
Pc08         TKFNCDEVKK-------                                            187
Pc10         KNVNCDDIKK-------                                            185
Pc11         KNIDCDDIKE-------                                            184
Pc12         KNVNCDDIKK-------                                            185
Pc14         KNVNCDDIKK-------                                            185
Pc16         KNVNCDDIKKKLVERST                                            192
Pc18         KNVNCDDIKK-------                                            185
Pc19         KNIDCDDIKE-------                                            184
Pc20         KKAKCDDDKK-------                                            188
Pc21         KNVNCENIQK-------                                            183
Pc22         KKAKCDDDKK-------                                            188
pallidipin 2 SKVNCDNMK--------                                            188
```

Fig. 1. (A) Sequence alignment of pallidipin 2-like proteins from *Panstrongylus chinai* (Pc02, Pc07, Pc08, Pc10-Pc12, Pc14, Pc16 and Pc18-Pc22) together with pallidipin 2 from *Triatoma pallidipennis* (accession number: AAA30329). Black-shaded amino acids represent identical amino acids and gray-shaded amino acids represent conserved amino acids. Dashes indicate gaps introduced for maximal alignment. Asterisks at the top of the amino acids denote conserved cysteine residues. (B) Phylogenetic tree analysis of pallidipin 2-like proteins (Pc02, Pc07, Pc08, Pc10-Pc12, Pc14, Pc16 and Pc18-Pc22) from *P. chinai* with pallidipin 2 from *T. pallidipennis*. The numbers in parentheses indicate the number of transcripts of each contig. The scale bar represents 0.1% divergence. Bootstrap values are shown above branches.

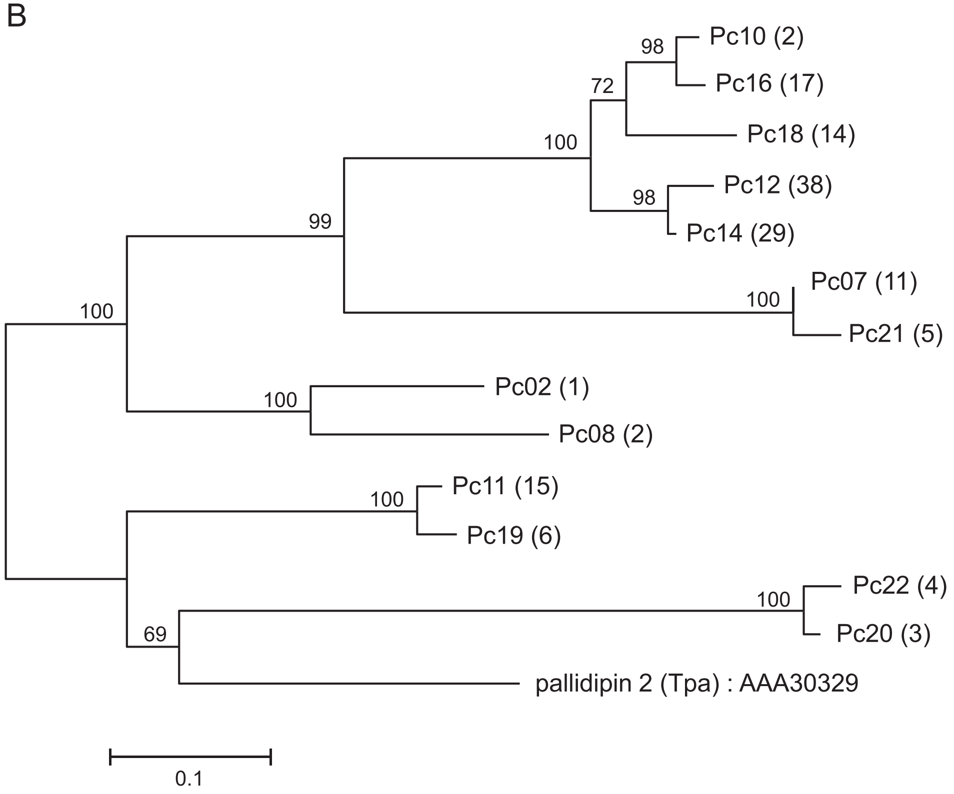
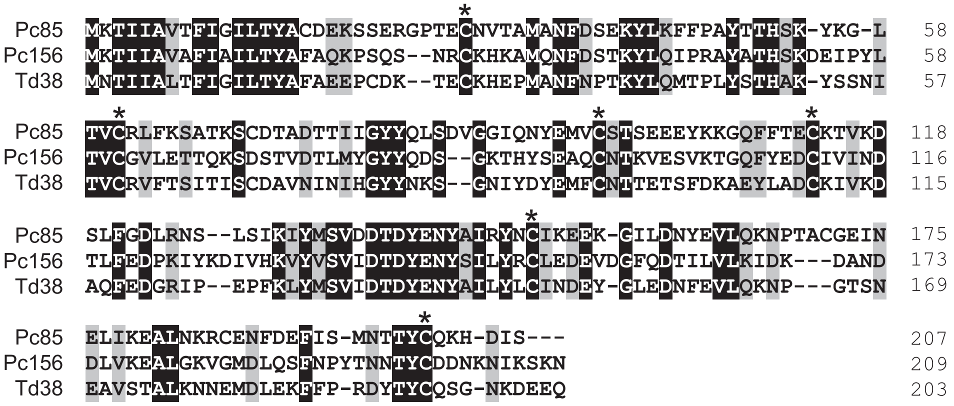


Fig. 1. (*continued*)



```
                                   *
Pc85   MKTIIAVTFIGILTYACDEKSSERGPTECNVTAMANFDSEKYLKFFPAYTTHSK-YKG-L   58
Pc156  MKTIIAVAFIAILTYAFAQKPSQS--NRCKHKAMQNFDSTKYLQIPRAYATHSKDEIPYL   58
Td38   MNTIIALTFIGILTYAFAEEPCDK--TECKHEPMANFNPTKYLQMTPLYSTHAK-YSSNI   57

         *                                   *               *
Pc85   TVCRLFKSATKSCDTADTTIIGYYQLSDVGGIQNYEMVCSTSEEEYKKGQFFTECKTVKD  118
Pc156  TVCGVLETTQKSDSTVDTLMYGYYQDS--GKTHYSEAQCNTKVESVKTGQFYEDCIVIND  116
Td38   TVCRVFTSITISCDAVNINIHGYYNKS--GNIYDYEMFCNTTETSFDKAEYLADCKIVKD  115

                                        *
Pc85   SLFGDLRNS--LSIKIYMSVDDTDYENYAIRYNCIKEEK-GILDNYEVLQKNPTACGEIN  175
Pc156  TLFEDPKIYKDIVHKVYVSVIDTDYENYSILYRCLEDEVDGFQDTILVLKIDK---DAND  173
Td38   AQFEDGRIP--EPFKLYMSVIDTDYENYAILYLCINDEY-GLEDNFEVLQKNP---GTSN  169

                                *
Pc85   ELIKEALNKRCENFDEFIS-MNTTYCQKH-DIS---                          207
Pc156  DLVKEALGKVGMDLQSFNPYTNNTYCDDNKNIKSKN                          209
Td38   EAVSTALKNNEMDLEKFFP-RDYTYCQSG-NKDEEQ                          203
```

Fig. 2. Sequence alignment of Td38-like proteins from *Panstrongylus chinai* (Pc85 and Pc156) together with Td38 from *Triatoma dimidiata* (accession number: AB470389). Black-shaded amino acids represent identical amino acids and gray-shaded amino acids represent conserved amino acids. Dashes indicate gaps introduced for maximal alignment. Asterisks at the top of the amino acids denote conserved cysteine residues.

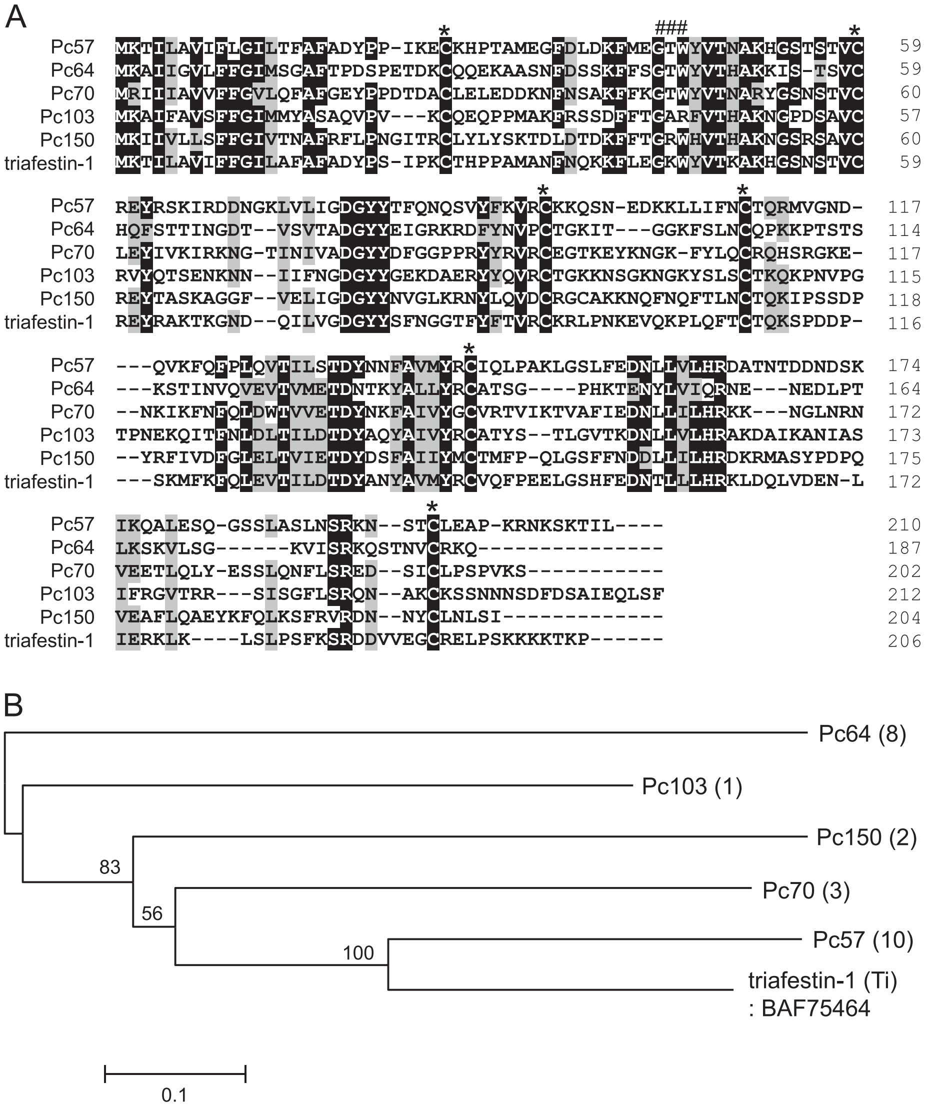


Fig. 3. (A) Sequence alignment of triafestin-1-like proteins from *Panstrongylus chinai* (Pc57, Pc64, Pc70, Pc103 and Pc150) together with triafestin-1 from *Triatoma infestans* (accession number: BAF75464). Black-shaded amino acids represent identical amino acids and gray-shaded amino acids represent conserved amino acids. Dashes indicate gaps introduced for maximal alignment. Asterisks at the top of the amino acids denote conserved cysteine residues, and the GXW motif is indicated by ###. (B) Phylogenetic tree analysis of triafestin-1-like proteins from *P. chinai* (Pc57, Pc64, Pc70, Pc103 and Pc150) with triafestin-1 from *T. infestans*. The numbers in parentheses indicate the number of transcripts of each contig. The scale bar represents 0.1% divergence. Bootstrap values are shown above branches.

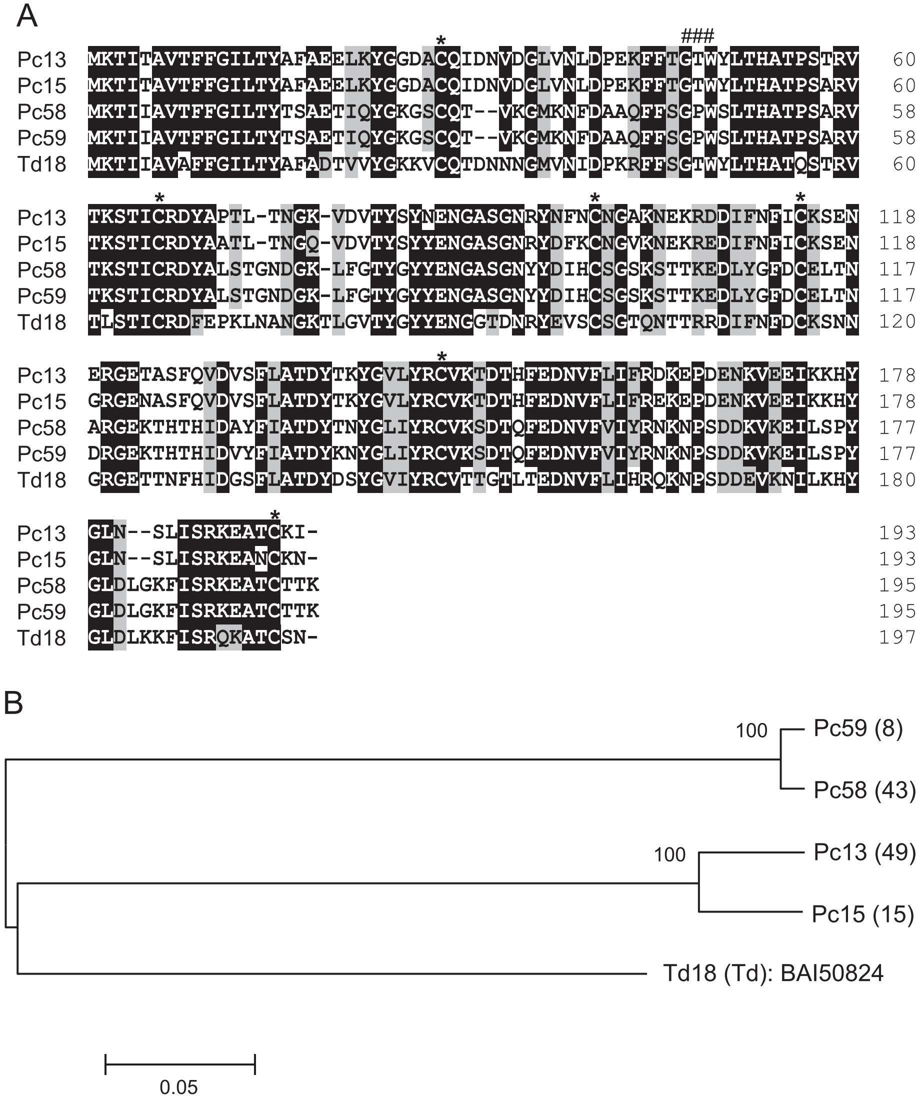


Fig. 4. (A) Sequence alignment of Td18-like proteins from *Panstrongylus chinai* (Pc13, Pc15, Pc58 and Pc59) together with Td18 from *Triatoma dimidiata* (accession number: BAI50824). Black-shaded amino acids represent identical amino acids and gray-shaded amino acids represent conserved amino acids. Dashes indicate gaps introduced for maximal alignment. Asterisks at the top of the amino acids denote conserved cysteine residues, and the GXW motif is indicated by ###. (B) Phylogenetic tree analysis of Td18-like proteins from *P. chinai* (Pc13, Pc15, Pc58 and Pc59) together with Td18 from *T. dimidiata*. The numbers in parentheses indicate the number of transcripts of each contig. The scale bar represents 0.05% divergence. Bootstrap values are shown above branches.

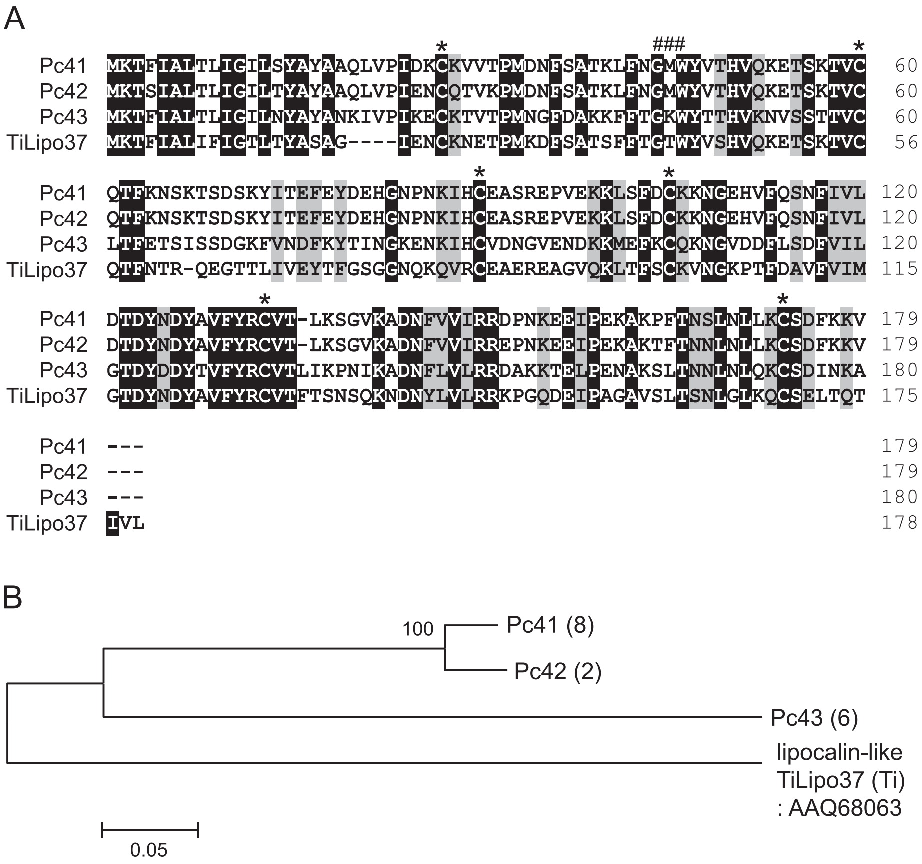


Fig. 5. (A) Sequence alignment of lipocalin-like TiLipo37-like proteins from *Panstrongylus chinai* (Pc41-Pc43) together with lipocalin-like TiLipo37 from *Triatoma infestans* (accession number: AAQ68063). Black-shaded amino acids represent identical amino acids and gray-shaded amino acids represent conserved amino acids. Dashes indicate gaps introduced for maximal alignment. Asterisks at the top of the amino acids denote conserved cysteine residues, and the GXW motif is indicated by ###. (B) Phylogenetic tree analysis of lipocalin-like TiLipo37-like proteins from *P. chinai* (Pc41-Pc43) together with lipocalin-like TiLipo37 from *T. infestans*. The numbers in parentheses indicate the number of transcripts of each contig. The scale bar represents 0.05% divergence. Bootstrap values are shown above branches.

Acknowledgements

We are grateful to Dr. José M. C. Ribeiro (Vector Biology Section, Laboratory of Malaria and Vector Research, NIAID, NIH, USA) for the development and training of all custom bioinformatics programs used in this research. This study was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (Grant nos. 25257501 and 15H04588).

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.09.039.

References

[1] H. Kato, R.C. Jochim, E.A. Gomez, S. Tsunekawa, J.G. Valenzuela, Y. Hashiguchi, Salivary gland transcripts of the kissing bug, *Panstrongylus chinai*, a vector of Chagas disease, Acta Trop. 174 (2017) 122–129.

[2] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[3] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.