

Supplemental data (Material and Methods)

2.6. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE)

Approximately 30 μg of venom was treated with LDS sample buffer (Invitrogen) containing SDS without reducing conditions and applied to a NU-PAGE 4-12% Bis-Tris gel (MES buffer) (Invitrogen) 1 mm thick. Protein bands were stained with 0.2% Coomassie Blue. For amino-terminal sequencing of proteins, 30 μg of venom was electrophoresed and transferred to polyvinylidene difluoride (PVDF) membrane using 10 mM CAPS, pH 11.0, 10% methanol as the transfer buffer on a Blot-Module for the XCell II Mini-Cell (Invitrogen). The membrane was stained with 0.02% Coomassie Blue in the absence of acetic acid. Stained bands were cut from the PVDF membrane and subjected to Edman degradation using a Procise sequencer (Perkin-Elmer Corp.). To estimate the molecular weight of the samples, SeeBlue™ markers from Invitrogen (myosin, BSA, glutamic dehydrogenase, alcohol dehydrogenase, carbonic anhydrase, myoglobin, lysozyme, aprotinin, and insulin, chain B) were used. To find the cDNA sequences corresponding to the amino acid sequence—obtained by Edman degradation of the proteins transferred to PVDF membranes from PAGE gels—we wrote a search program (in VB) that checked these amino acid sequences against the three possible protein translations of each cDNA sequence obtained in the mass sequencing project of *B. gabonica* venom gland. This program was written using the same approach utilized in the BLOCKS (Henikoff and Henikoff, 1994) or PROSITE (Sibbald et al., 1991) databases. The program is very useful when mixed sequence information occurs, for example, amino terminal sequences deriving from a mix of equal peptides. In this case, two different cDNA sequences may be unambiguously found as matches.

2.7. Snake venom gland cDNA library construction

A fragment was rapidly obtained from the center part of the gland. Fragments were transferred to a sterile plastic Petri dish located on the top of dry ice to avoid melting. *B. gabonica* salivary gland mRNA was obtained using Micro-Fast Track mRNA isolation kit (Invitrogen, San Diego, CA, USA) according to the manufacturer's instructions. The PCR-based

cDNA library was made following the instructions for the SMART cDNA library construction kit (Clontech, Palo Alto, CA, USA) as described (Francischetti et al., 2002). *B. gabonica* venom gland mRNA (~200 ng) was reverse transcribed to cDNA using Superscript II RNase H-reverse transcriptase (Gibco-BRL, Gaithersburg, MD, USA) and the CDS III/3' PCR primer (Clontech) for 1 h at 42 °C. Second-strand synthesis was performed through a PCR-based protocol using SMART III (Clontech) as the sense primer and CDS III/3' as the antisense primer. These two primers create *Sfi*IA and B sites at the ends of the nascent cDNA. Double-strand (ds) cDNA synthesis was performed on a Perkin Elmer 9700 thermacycler (Perkin Elmer Corp., Foster City, CA, USA) using Advantage Klen-*Taq* DNA polymerase (Clontech). PCR conditions were as follows: 94 °C for 2 min; 19 cycles of 94 °C for 10 s and 68 °C for 6 min. ds cDNA was immediately treated with proteinase K (0.8 µg/µl) for 20 min at 45 °C and washed three times with water using Amicon filters with a 100-kDa cutoff (Millipore). ds cDNA was then digested with *Sfi*I for 2 h at 50 °C. The cDNA was then fractionated using columns provided by the manufacturer (Clontech). Fractions containing cDNA were pooled as large-, medium- and small-sized cDNA, concentrated, and washed three times with water using an Amicon filter with a 100-kDa cutoff. For the large-, medium- and small-sized fractions, cDNA were separately combined and concentrated to a volume of 10 µl. The concentrated cDNA were then ligated overnight at 16 °C into a Lambda TriplEx2 vector (Clontech), and the resulting ligation reaction was packed using Gigapack Gold III from Stratagene/Biocrest (Cedar Creek, TN, USA) following manufacturer's specifications. The three libraries (large-, medium- and small-sized) obtained were plated by infecting log-phase XL1-Blue cells (Clontech) at appropriate dilutions and the amount of recombinants determined by PCR using vector primers flanking the inserted cDNA and visualized on a 1.1% agarose gel with ethidium bromide (EtdBr) (1.5 µg/ml).

2.8. Sequencing of *B. gabonica* cDNA library

The *B. gabonica* salivary gland cDNA library was plated to approximately 200 plaques per plate (150-mm Petri dish). The plaques were randomly selected and transferred to a 96-well polypropylene plate containing 100 μ l of water per well. The plate was covered and placed on a gyratory shaker for 1 h at room temperature. The phage sample (5 μ l) was used as a template for a PCR reaction to amplify random cDNA. The primers used for this reaction were sequences from the TriplEx2 vector and were named PT2F1 (5'-AAG TAC TCT AGC AAT TGT GAG C-3'), which is positioned upstream of the cDNA of interest (5' end), and PT2R1 (5'-CTC TTC GCT ATT ACG CCA GCT G-3'), which is positioned downstream of the cDNA of interest (3' end). Platinum *Taq* polymerase (Gibco-BRL) was used for these reactions. Amplification conditions were: 1 hold at 75 °C for 3 min, 1 hold at 94 °C for 2 min, and 33 cycles at 94 °C for 1 min, 49 °C for 1 min, and 72 °C for 1 min 20 s. Amplified products were visualized on a 1.1% agarose gel with EtdBr. The concentration of ds cDNA was measured using Hoechst dye 33258 on a Fluorolite 1000 plate fluorometer (Dynatech Laboratories, Chantilly, VA, USA). PCR reactions (3-4 μ l) containing between 100 and 200 ng of DNA were then treated with Exonuclease I (0.5 units μ l⁻¹) and shrimp alkaline phosphatase (0.1 units μ l⁻¹) for 15 min at 37 °C and 15 min at 80 °C on a 96-well PCR plate. This mixture was used as a template for a cycle sequencing reaction using the DTCS labeling kit from Beckman Coulter Inc. (Fullerton, CA, USA). The primer used for sequencing (PT2F3) is upstream of the inserted cDNA and downstream of primer PT2F1. Sequencing reaction was performed on a Perkin Elmer 9700 thermocycler. Conditions were 75 °C for 2 min, 94 °C for 4 min, and 30 cycles of 96 °C for 20 s, 50 °C for 20 s, and 60 °C for 4 min. After cycle sequencing the samples, a cleaning step was done using the multiscreen 96-well plate cleaning system (Millipore). The 96-well multiscreening plate was prepared by adding a fixed amount (manufacturer's specification) of Sephadex-50 (Amersham Pharmacia Biotech, Piscataway, NJ, USA) and 300 μ l of deionized water. After 1 h of incubation at room temperature, the water was removed from the multiscreen plate by centrifugation at 750 g for 5 min. After the Sephadex in the multiscreen plate was partially dried, the whole cycle-sequencing reaction was added to the center of each well,

centrifuged at 750 g for 5 min, and the clean sample was collected on a sequencing microtiter plate (Beckman Coulter Inc.). The plate was then dried on a Speed-Vac SC 110 model with a microtiter plate holder (Savant Instruments Inc., Holbrook, NY, USA). The dried samples were immediately resuspended with 25 μ l of deionized ultrapure formamide (J. T. Baker, Phillipsburg, NJ, USA), and one drop of mineral oil was added to the top of each sample. Samples were either sequenced immediately on a CEQ 2000 DNA sequencing instrument (Beckman Coulter Inc.) or stored at -30°C.

2.9. Sequence information cleaning

Raw sequences originating from the DNA sequencer had either of five letters in their result: ATCG for the nucleotide bases, and N when the sequencer program could not call a base. Usually the beginnings and ends of sequences have a higher proportion of N calls. Sequences also contain primer and vector sequences used in library construction. For this reason, raw sequences were treated by a program written in VisualBasic 6.0 (VB) (Microsoft Corp., Redmond, WA, USA) as follows: *i*) Sequences were analyzed in their first 80 base pairs (bp) for groups of four Ns, and, if found, the block of four Ns closer to position 80 was used to trim the raw sequence from this 5' N-rich region. *ii*) For sequences longer than 110 bp, windows of 10 bp were screened for the occurrence of four or more Ns above position 100. The positive window with the smallest position value was used to trim the sequence from the 3' N-rich region. Sequences thus trimmed and having more than 10% N content were discarded. *iii*) Good-quality and trimmed sequences were then searched for occurrence of the primers used in library construction (the SMART III primer as well as the CDS/R primers). A moving window the size of the primer was searched on the sequence for matches with the primer sequence. If more than a 70% match was obtained, or if a contiguous match longer than 50% of the length of the primer was observed, the sequence was trimmed at the beginning or end of the window depending on the expected position of the primer. This simple algorithm avoided errors due to spurious insertions. *iv*) The trimmed sequence was "polished" by removing any trailing N. The sequence final N content was assessed, as well as its AT content and length. The final sequence was written to a FASTA-format file containing in its definition line the actions taken by the program.

2.10. Sequence clustering

The FASTA file containing all sequences was clustered by first blasting each sequence against the formatted database file using Blastn with the filter option off (-FF). The blast output was ccanned and sequences having > 92% identity over 100 pb were joined in clusters. These were further treated by the program CAP assembler (Huang, 1992) to produce contigs of related sequences. Finally, a program was written in VB that combines all the results to create Table 1 of this paper, except for the "Comments" column. The output of this program is imported into a Microsoft Excel spreadsheet. In the supplemental material, Table 1 includes hyperlinks to the best NR protein match in the NCBI site, all FASTA files for each individual cluster, CAP assembler alignments files for each cluster, when available, and the FASTA file for the whole database. Each cluster was individually analyzed for the probable function of its translation product and assigned a probably secreted, probably housekeeping, or indeterminate function. This decision was based on the best match to the NR protein database and related sequences as searched online at the NCBI site (www.ncbi.nlm.gov) and on the SMART and/or Pfam matches, including searches of the nature of the domains by online searches of the respective sites.

2.11. Searches for known sequence similarities and known protein domains of the cDNA sequences

To obtain information on the possible role of the cDNA sequences, the FASTA file containing all the stripped sequences was blasted against the GenBank nonredundant (NR) protein database from the National Center for Biotechnology Information (NCBI) using the stand-alone BlastX program found in the executable package at <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/> (Altschul et al., 1997). The NR database as well as the cumulative updates were regularly downloaded, uncompressed with GUNZIP (found at www.gzip.org/), and formatted for Blast program use with the FORMATDB program (executables also found at <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/>) with the help of a program written in PERL code (software found at www.activeperl.com). NCBI sequences are indicated in this manuscript by their accession number as gi|XXXX where XXXX is a unique identifier number. The resulting file was parsed, and the best match was incorporated in the FASTA definition line

after the delimiter |. The sequences were next submitted to the stand-alone program RPSBlast (Altschul et al., 1997) and searched against the Conserved Domains Database (found at <ftp://ftp.ncbi.nlm.nih.gov/pub/mmdb/cdd/>), which includes all Pfam (Bateman et al., 2000) and SMART (Schultz et al., 2000) protein domains. The RPSBlast result file was parsed as above and the best match incorporated also into the FASTA definition line of the sequence. When all sequences of a particular cluster were blasted against the NR protein database (using the BlastX program), the best protein match was searched for the species from which the NR database sequence originated. If the species was not *B. gabonica* or no matches to the NR were found, the cluster was marked as representing a novel *B. gabonica* sequence. All cluster sequences that gave a match to a *B. gabonica* protein sequence were further individually inspected to verify whether the cDNA sequence represented nearly the same information translated as the protein match or a closely related but different protein.

2.12. Full-length sequencing of selected cDNA clones

An aliquot (4 μ l) of the lambda phage containing the cDNA of interest was amplified using the PT2F1 and PT2R1 primers (conditions as described above). The PCR samples were cleaned using the multiscreen-PCR 96-well filtration system (Millipore). Cleaned samples were sequenced first with PT2F3 primer and subsequently with custom primers. Primer selection for complete sequence of selected full-length cDNA was also assisted by a program (written in VB) that identified unique primer sites within the sequences. To assemble the sequences, the previously known sequence was assembled with the new sequence using the CAP assembler (Huang, 1992). The program attempted to locate a polyA region by using a 20-bp window in which 19 As would constitute a polyA. If no polyA was found, a new set of primers would be found to continue extension of the cDNA. The program also produces a consensus output and the three possible translations of this unidirectionally cloned RNA. The final alignment is adjusted by hand. If necessary, the original tracings of the DNA sequencer are reviewed for critical base calls. The translated sequences are submitted as a FASTA file to the SIGNALP server (at <http://www.cbs.dtu.dk/services/SignalP/>) (Nielsen et al., 1997), which responds by e-mail indicating whether a signal peptide exists and its location. A program written in VB interprets this SIGNALP result file and removes the signal peptide, if it is predicted to exist, to

create a mature protein sequence. Molecular weights using average molecular masses for C, H, O, N, P, and S are calculated for all protein sequences, as are isoelectric points (pI) based on reduced proteins, following the pKa for amino acids within proteins as indicated before (Altland, 1990; Bjellqvist et al., 1994). This program produced an output that can be read by the spreadsheet program Excel to produce Table 1 in this paper. In the supplemental table, hyperlinks are given to all proteins. Alignments of protein sequences were done with the ClustalW program version 1.7 (Thompson et al., 1994). Dendrograms were obtained using ClustalX version 1.8, and the output was used to generate filogenetic trees using TreeView software (Page, 1996). The complete list of the sequences coding for proteins with secretory, housekeeping, or undetermined function, with or without database hits, can be obtained on request (ifrancischetti@niaid.nih.gov); the spreadsheet includes: *i*) columns with hyperlinks to the best match of the NR database, *ii*) links to NR matches found for the cluster, *iii*) matches to the Conserved Domains Database, *iv*) FASTA-formatted files for each cluster, and *v*) CAP-assembler alignments of each contig having two or more sequences.

2.12. Statistical tests

Statistical tests were performed with SigmaStat version 2.0 (Jandel Software, San Rafael, CA, USA). Kruskal-Wallis ANOVA on ranks was performed, and multiple comparisons were made by the Dunn method. Dual comparisons were made with the Mann-Whitney rank sum test.

Supplemental data (References)

- Altland, K., 1990. IPGMAKER: a program for IBM-compatible personal computers to create and test recipes for immobilized pH gradients. *Electrophoresis* 11, 140-147.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389-3402.
- Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., Sonnhammer, E.L., 2000. The Pfam protein families database. *Nucleic Acids Res.* 28:263-266.
- Bjellqvist, B., Basse, B., Olsen, E., Celis, J.E., 1994. Reference points for comparison of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 15, 529-539.
- Henikoff, S., Henikoff, J.G., 1994. Protein family classification based on searching a database of blocks. *Genomics* 19, 97-107.
- Higgins, D.G., Thompson, J.D., Gibson, T.J., 1996. Using CLUSTAL or multiple sequence alignments. *Methods Enzymol.* 266, 383-402.
- Huang, X., 1992. A contig assembly program based on sensitive detection of fragment overlaps. *Genomics* 14, 129-149.
- Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1-6.
- Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Appl. Biosci.* 12, 357-358.
- Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., Bork, P., 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucl. Acids Res.* 28, 231-234.
- Sibbald, P.R., Sommerfeldt, H., Argos, P., 1991. Automated protein sequence pattern handling and PROSITE searching. *Comput. Appl. Biosci.* 7, 535-536.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22, 4673-4680.

Supplemental data (Legends to Figures)

Fig. 6. (A) Alignment of *B. gabonica* serine proteases and (B) unrooted cladogram. See Figure 3 legend for an explanation of the layout. BG_SP (*B. gabonica* serine protease; AY 430410); ML_SP (*M. lebetina* serine protease; gi 22417221); TF_SP (*T. flavoviridis* serine protease; gi 3915685); BA_SP (*B. atrox* serine protease; gi114837); BJ_SP (*B. jararaca* serine protease; gi 13959657); AB_SP (*A. blomhoffi* serine protease; gi 6706013); and GU_SP (*G. ussuriensis* serine protease; gi 17933276). The catalytic triad amino acids are boxed.

(A)

```

BG_Kunitz_1_  MSSSGLLLLLGLLTLWAE LTPVSGKNRPEFCNLPADTGFCKAYEPRFYDVSKECCQKFT
BG_Kunitz_2_  MSSSGLLLLLGLLTLWAE LTPVSGKKRPDEFCYLPADTGFCKMANFPRFYDSASKKCKKFT
PT_Kunitz_   MSSSGLLLLLGLLTLWEV LTPVSSKDHPKFCLELPADTGSCKGNVPRFYFNADHHCCQKFI
VA_Kunitz_   -----RDRPKFCYLPADPGRC LAYMPRFYFNPA SNKCKE KFI
BF_Kunitz_   -----KNRPTFCNLLPETGRONALI PAFYFN SHLHKCKQKFN
HH_Kunitz_   -----RPDFCELPAETGLCKAYI RSFHYNLAQQCLQFI
OS_Kunitz_   -----KDRPKFC HLPKPGFCRAAI PRFYFNPHSKCKE KFI
EM_Kunitz_   -----FCYLPDDPGVCKAHI PRFYFNPA SNKCKNFI
DP_Kunitz_   --SGHLLLLLGLLTLWAE LTPVSG--AAKYCKLPLRIGCKRKI PSFYKWKAKCCLPFD
                                     : * * * * * * : * *

```

```

BG_Kunitz_1_  YGGCKGNSNNEFSMDECRKTCV ASATRRPT-
BG_Kunitz_2_  YGGCHGNANNEETREECRKTCF ASAARRPT-
PT_Kunitz_   YGGCGGNANNEFKTIEECKSTCAA-----
VA_Kunitz_   YGGCRGNANNEFKTWDECRHTCVASG-IQPR-
BF_Kunitz_   YGGCGGNANNEFKTIDEQRTC AAKYGRSS--
HH_Kunitz_   YGGCGGNANREFKTIDEQRTCV G-----
OS_Kunitz_   YGGCHGNANKFKTPDECRNYTCLGVSL-----
EM_Kunitz_   YGGCGGNANNEETRAEQRHTCVASGKGGPRP
DP_Kunitz_   YSGCGGNANREFKTIEEQRTCV G-----
* . * * * * : * * : * * .

```

(B)

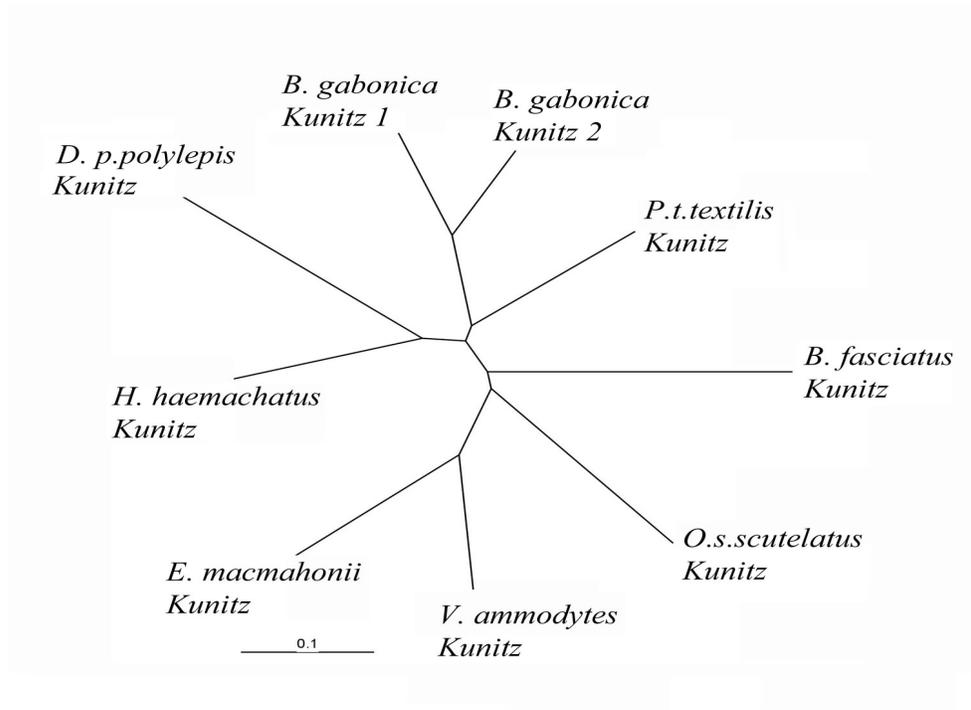


Figure 7

Fig. 7. (A) Alignment of *B. gabonica* Kunitz inhibitors and (B) unrooted cladogram. See Figure 3 legend for an explanation of the layout. The bar represents the degree of divergence among sequences. BG_Kunitz_1 (*B. gabonica* Kunitz inhibitor 1 or Bitisilin-1, AY430402), BG_Kunitz_2 (*B. gabonica* Kunitz inhibitor 2 or Bitisilin-2, AY430413); PN_Kunitz (*Pseudonaja textilis textilis* Kunitz inhibitor, gi 15321636); VA_Kunitz (*Vipera ammodytes ammodytes* Kunitz inhibitor, gi 125043); BF_Kunitz (*Bungarus fasciatus* Kunitz inhibitor, gi 125033); HH_Kunitz (*Hemachatus haemachatus* Kunitz inhibitor, gi 125039); OS_Kunitz (*Oxyuranus s. scuttelatus* Kunitz, gi 348615); EM_Kunitz (*Eristocophis macmahonii* Kunitz inhibitor, gi 125049); and DP_Kunitz (*Dendroaspis p. polylepis* Kunitz inhibitor, gi 25754451). The conserved cysteines are boxed.

(A)

```

BG_Cystatin_  -----MHSRLLVAAPHCLLLLLLLPSALPAL-KVGGLYPRDVMDEPVEAAAFAVENYN
BA_Cystatin_  -----I P-----GGLSPRDVTDPDVQEAFAFAVEKYN
NA_Cystatin_  -----I P-----GGLSPRSVSDPDVQKAAAFVQEYN
TF_Cystatin_  -----MHSRLLVAAPHCLLLLLLLPSALPAL-KVGGLYPRDVMDEPVEAAAFAVENYN
EW_Cystatin_  --MAGARGCVLLAAALMLVGAVLGSEDRSR--LLGAPVPVDENDEGLQRALQFAMAEYN
HS_Cystatin_  MARSNLPALGLLALVAFCLLALPRDARARQPQERMVGLRDLSPDDPQVQKAAQAAVASYN
                .      *      .      *      :*. *      * : .**

BG_Cystatin_  AQRSTNDNYFKARRIVEAQSQVSVGVKYYLKMELAKTTCKKIAGKPKLYQEIQNCNLPPEN
BA_Cystatin_  AGSKNDYYFKERRVVEAQSQVSVGVKYYLMMELKTTCKKTVGRPKGYQEIQNCNLPPEN
NA_Cystatin_  AGSANAHYKELRVVEAQSQSVAGEKYFLMMELVKTKCAKTAGKPKVYKEIQNCELPPIK
TF_Cystatin_  AQRSTNDNYFKARRIVEAQSQVSVGVKYYLKMELAKTTCKKIAGKPKLYQEIQNCNLPPEN
EW_Cystatin_  RASNDKYSSRVVVISAKRQLVSGIKYILQVEIGRTTCFKSSG-----DLQSCEFHDEP
HS_Cystatin_  MGSNSIYYFRDTHIIKAQSQLVAGIKYFLTMEMGSTDCRKRTRVTD-HVDLTTCPAAGA
                * .      :      :.:. * : * * : * * * : * : * * *      : : . * :

BG_Cystatin_  QQEEI-TCHFVWSRPWLQKTVLTKDEL---
BA_Cystatin_  QQEEI-TCRFEVWSRPWLPSTSLTK-----
NA_Cystatin_  QQEEK-LCGFQVW-----
TF_Cystatin_  QQEEI-TCHFVWSRPWLQKTVLTKDEL---
EW_Cystatin_  EMAYKTTCTFVVYSIPWLNQIKLLESKCQ--
HS_Cystatin_  QQEKL-RCDFEVLVVPWQNSQLLKHNCVQM
                :      :      * * *
    
```

(B)

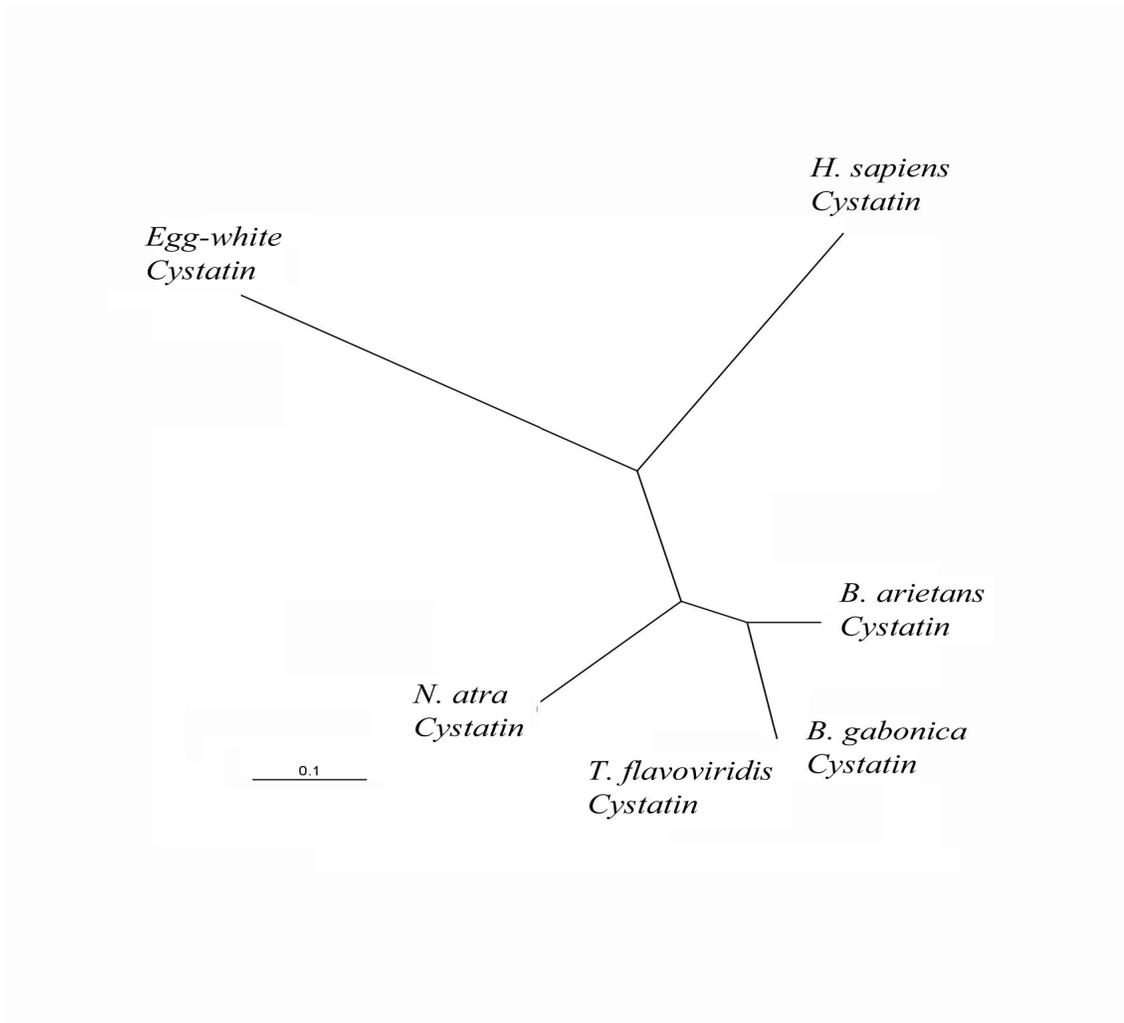


Figure 8

Fig. 8. (A) Alignment of *B. gabonica* cystatin inhibitors and (B) unrooted cladogram. See Figure 3 legend for an explanation of the layout. The bar represents the degree of divergence among sequences. BG_cystatin (*Bitis gabonica* cystatin or Biticystatin, AY430403); BA_cystatin (*Bitis arietans* cystatin, gi 118194); NA_cystatin (*Naja atra* cystatin, gi 18202373); TF_cystatin (*Trimerusurus flavoviridis* cystatin, gi 321083); EW_cystatin (egg-white cystatin, gi 118195); and HS_cystatin (*Homo sapiens* cystatin, gi 453113).

Fig. 9. (A) Alignment of *B. gabonica* C-type lectins and (B) unrooted cladogram. See Figure 3 legend for an explanation of the layout. The bar represents the degree of divergence among sequences. BG_Lectin_1 (*B. gabonica* C-type lectin-1; gi AY439477); BG_Lectin_2 (*B. gabonica* C-type lectin-2; gi AY429478); BG_Lectin_3 (*B. gabonica* C-type lectin-3; gi AY429479); GH_Lectin (*G. halys* C-type lectin; gi 4337050); DA_Lectin (*D. acutus* C-type lectin; gi 23321261); TS_Lectin (*T. stejnegeri* C-type lectin; gi 7674107); TF_Lectin (*T. flavoviridis* C-type lectin; gi 2851435); BJ_Lectin (*B. jararaca* lectin; gi 37537732); and BF_Lectin (*B. fasciatus* lectin; gi 13876735).

(A)

```

BG_PLA2_ MRALWIVAVWLMGVEGHLEQFGNMI DHVSGR-SFWLYVSYGICYCGWGGSGKPDATDRCC
VP_PLA2_ MRTLWIIVAVCLMGVEGHILTQFGDMINKKTGT FGLLSYVYGYCYCGLGGKGPQDATDRCC
EP_PLA2_ MRTLWIIVAVWLMGVEGNLYQFGKMIKNKTGKPAMFSYSAYGICYCGWGGQKPDPSDRCC
TF_PLA2_ MRTLWIMAVLLVGVKGHLMQFENMIKKVTGRSGIWWYGSYGYCYCGKGGEGRPQDPSDRCC
CV_PLA2_ MRTLWILAVLLLGVEGNLVQFELLIMKVAKRSGLLSY SAYGICYCGWGGYGRPQDATDRCC
BI_PLA2_ MRTLWIMAVLLVGVGNLWQFGKMMNYVMGQSVVYKY FYYGICYCGWGGIGQPRDATDRCC
*:***:* * :***:*** ** :: . * ***** ** * :* :***

BG_PLA2_ FVHDCCYGKMGTYNTKWTSYNYEFQNGDI ICGDEDPRKKELCECDRVA AICFGNNRNTYN
VP_PLA2_ FVHDCCYGTVNGCDEKLSYYSYFQNGDIVCGDDDPCLRAVCECDRVA AICFGENMNTYD
EP_PLA2_ FMHDCCYTRVNNCSEKMTLYSYRFENGDI ICGDNDPCRKAVCECDREA AICLGENVNTYD
TF_PLA2_ FVHDCCYGKVTGCDEKIDDFYIYSSENGDIVCGDDDLCKKEVCECDKAA AICFRDNMDTYQ
CV_PLA2_ FVHDCCYGKVTDCNEKKTASYTYSENGEIVCGDDDPCKKQVCECDRVA AICFRDNIPSYD
BI_PLA2_ FVHDCCYGKVTGCDEKIDSYTYSKENGDVVC GDDDPCKQICECDRVAATCFRDNKDTYD
*:***** : . . * * * :***:***. : :***** : * * : * :**

BG_PLA2_ SKYFAYSSTNCMEE-TEQC
VP_PLA2_ KKYMLYSFFDCMEE-SEKC
EP_PLA2_ EKYRFYSSSYCTEESEKC
TF_PLA2_ NKYWFYPASNCKEE-SEPC
CV_PLA2_ NKYIQFPAKNCQEK-PEPC
BI_PLA2_ MKYWLYGAKNCQEE-SEPC
** : * * : . * *
    
```

(B)

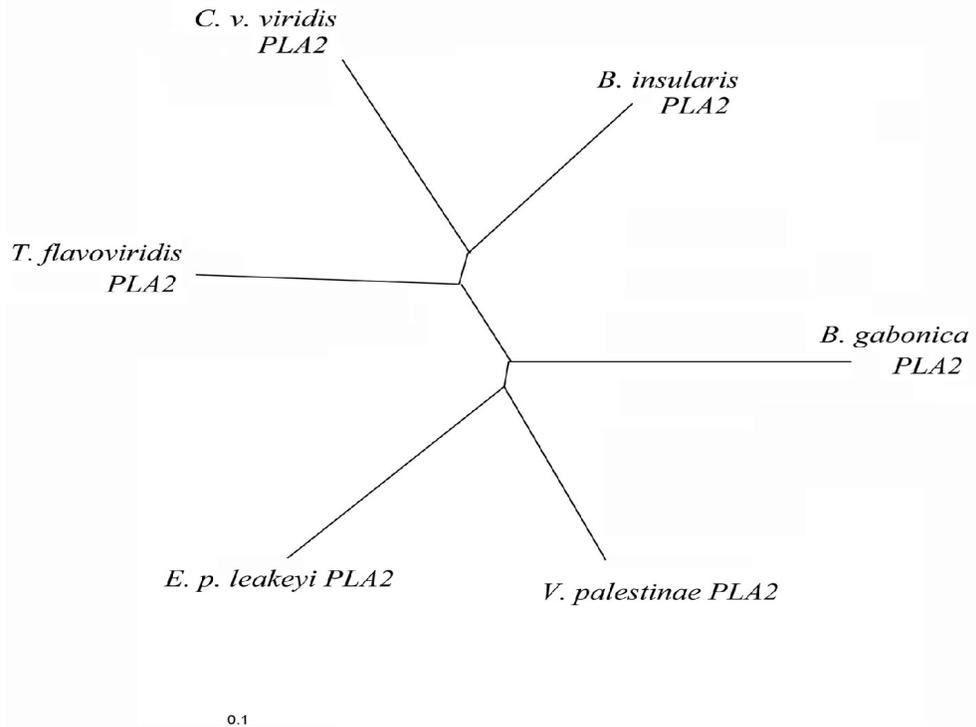


Figure 10

Fig. 10. (A) Alignment of *B. gabonica* PLA2 and (B) unrooted cladogram. See Figure 3 legend for an explanation of the layout. BG_PLA2 (*Bitis gabonica* PLA2, AY 429476); VP_PLA2 (*Vipera palestinae* PLA2, gi 6647690); EP_PLA2 (*Echis pyramidum leakeyi* PLA2, GI 27734438); TF_PLA2 (*Trimerusus flavoviridis* PLA2, gi 28201850); CV_PLA2 (*Crotalus viridis viridis* PLA2, gi 28893826); and BI_PLA2 (*B. insularis* PLA2, gi 26006832).

(A)

```

BG_VEGF_ MNFLLTWIHWGLAALLYFHNAKVLQAAPAQGDGERQQGEVIFFLKVYERSICRPVETMVD
GG_VEGF_ MNFLLTWIHWGLAALLYLQSAELSKAAPALGDGERKFNEVIKFLLEVYERSFCRTIETLVD
HS_VEGF_ MNFLLSWVHWSLALLLYLHHAKWSQAAPMAEGGGQNHHEVVKFMDVYQRSYCHPIETLVD
BI_VEGF_ MAAYL-----LAVAILFC----IQGWPL---G-TVQGQVMPFMEVYRHSVCQTRETLVS
BJ_VEGF_ MAVYL-----LAVAILFC----IQGWPL---G-TVQGQVMPFMEVYRHSVCQPRETLVS
*   *           **   :   :           :. *   *           :*: **:***:* *:. **:*.

BG_VEGF_ IFQEYPDEVEYIFKPSCVPLMRCAGCCGDEGLECVPDVYNVTMEIARIKPHQS-QHIAH
GG_VEGF_ IFQEYPDEVEYIFRPSCVPLMRCAGCCGDEGLECVPDVYNVTMEIARIKPHQS-QHIAH
HS_VEGF_ IFQEYPDEIEYIFKPSCVPLMRCAGCCGDEGLECVPTESNITMQIMRIKPHQG-QHIGE
BI_VEGF_ ILEEHDPDEVSHIFRPSCVTALRCGGCCTDESLKCTATGKR SVGREIMRVDPHKGT SKTEV
BJ_VEGF_ ILEEYPGEISHIFRPSCVTALRCGGCCTDESLECTATGKR SVGREIMRSLPHKGTSEKEV
*::*:*.***:***:***. :***.*** **.*:***. .: : : :*.***. .

BG_VEGF_ VSFQQHSHKCECRPKKDIRNK-DNHCEPCSERRKHLYKQDPLTCKCSCKAPDLRCKSKQLE
GG_VEGF_ MSFLQHSKDCRPPKDVKNKQENHCEPCSERRKHLFVQDPQTCKCSCKFTDSRCKSRQLE
HS_VEGF_ MSFLQHNKCECRPKKDRARQ-ENPCGPCSERRKHLFVQDPQTCKCSCKNTDSRCKARQLE
BI_VEGF_ MQFTEHTDCECRPRSASGVN-----S--RKH--KRNPEEGEPRAKFFPV-----
BJ_VEGF_ MQFTEHTDCECRPRSASGVN-----S--RKH--KRNPEEGEPRAKFFPV-----
:. * :*.***:***. :           *   **   :*: :   :   :*.

BG_VEGF_ LNERTCR CERPRR
GG_VEGF_ LNERTCRCEKPRR
HS_VEGF_ LNERTCRCDKPRR
BI_VEGF_ -----
BJ_VEGF_ -----
    
```

(B)

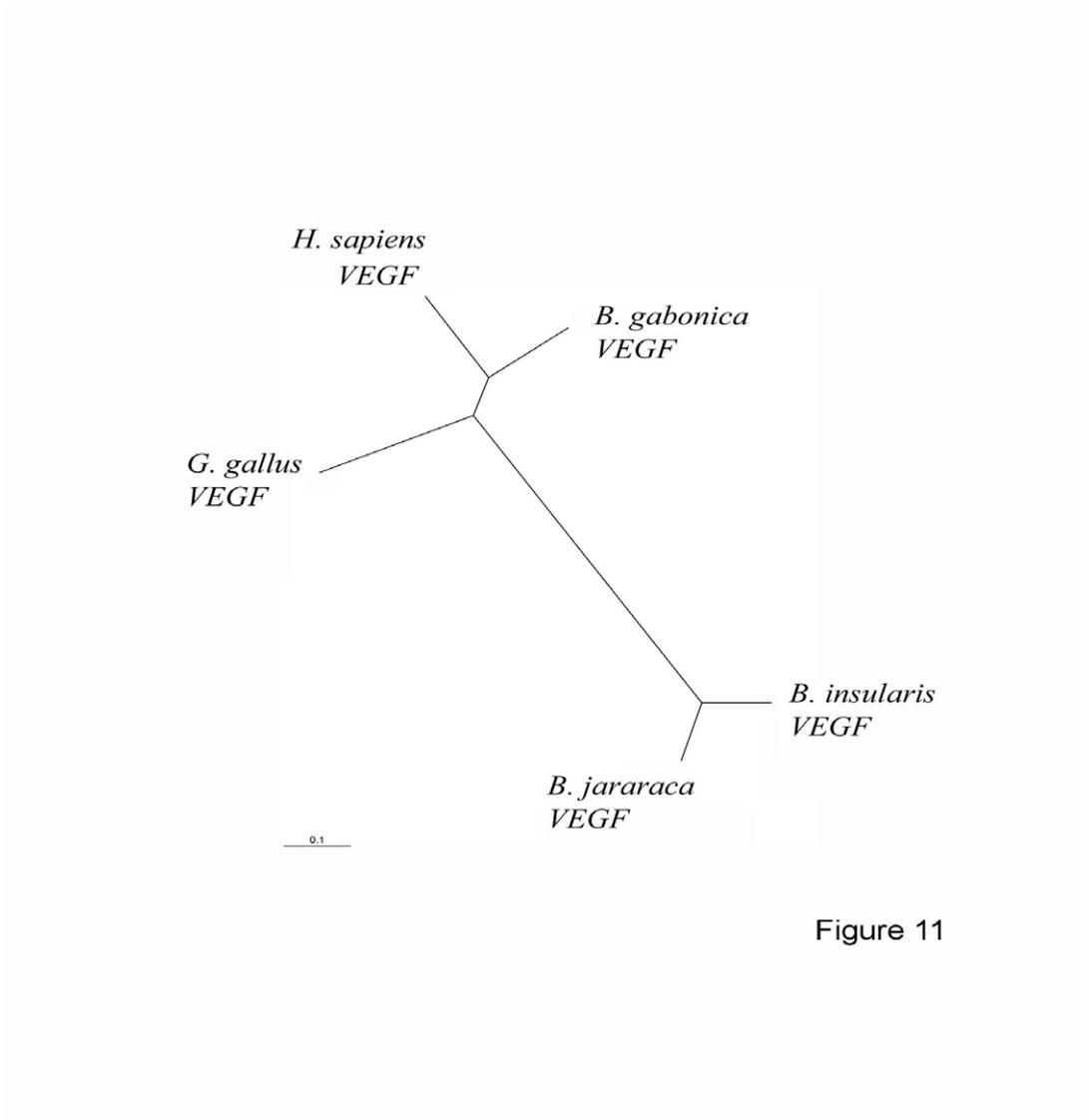


Figure 11

Fig. 11. (A) Alignment of *B. gabonica* VEGF and (B) unrooted cladogram. See Figure 3 legend for an explanation of the layout. The bar represents the degree of divergence among sequences. BG_VEGF (*B. gabonica* VEGF, AY429481); GG_VEGF (*Gallus gallus* VEGF, gi 27368068); HS_VEGF (*Homo sapiens* VEGF, gi 37659); BI_VEGF (*Bothrops insularis* VEGF, gi 15072460); BJ_VEGF (*Bothrops jararaca* VEGF, gi 15072460).

Supplemental data (Figures) – see next page