ELSEVIER

Toxicon 45 (2005) 547–560

TOXICON

www.elsevier.com/locate/toxicon

Sphingomyelinase D from venoms of *Loxosceles* spiders: evolutionary insights from cDNA sequences and gene structure☆

Greta J. Binford*, Matthew H.J. Cordes, Michael A. Wells

Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, Az 85719, USA

Received 2 August 2004; accepted 1 November 2004

Abstract

Loxosceles spider venoms cause dermonecrosis in mammalian tissues. The toxin sphingomyelinase D (SMaseD) is a sufficient causative agent in lesion formation and is only known in these spiders and a few pathogenic bacteria. Similarities between spider and bacterial SMaseD in molecular weights, p*I*s and N-terminal amino acid sequence suggest an evolutionary relationship between these molecules. We report three cDNA sequences from venom-expressed mRNAs, analyses of amino acid sequences, and partial characterization of gene structure of SMaseD homologs from *Loxosceles arizonica* with the goal of better understanding the evolution of this toxin. Sequence analyses indicate SMaseD is a single domain protein and a divergent member of the ubitiquous, broadly conserved glycerophosphoryl diester phosphodiesterase family (GDPD). Bacterial SMaseDs are not identifiable as homologs of spider SMaseD or GDPD family members. Amino acid sequence similarities do not afford clear distinction between independent origin of toxic SMaseD activity in spiders and bacteria and origin in one lineage by *ancient* horizontal transfer from the other. The SMaseD genes span at least 6500 bp and contain at least 5 introns. Together, these data indicate *L. arizonica* SMaseD has been evolving within a eukaryotic genome for a long time ruling out origin by *recent* transfer from bacteria.

Keywords: Venom; *Loxosceles*; *Corynebacteria*; Horizontal transfer; Concerted evolution

1. Introduction

The evolution of novel gene function is of fundamental interest across broad disciplines, particularly when the novel effect is toxic to humans. Venoms of predatory animals are rich in toxins with novel structure and function, and thus provide an arena for studies of molecular diversification (see Menez, 2002, for recent examples). These studies also provide valuable insight into taxonomic distribution and variation in venom toxins that have damaging or lethal effects on humans, information that is critical for developing broadly effective antibody-based diagnostics and treatments of envenomation.

Venoms from spiders of the genus *Loxosceles*, brown or violin spiders, are notorious for their ability to induce dermonecrotic lesions in mammalian tissues. The venom toxin sphingomyelinase D (SMaseD) is a sufficient causative agent for lesion formation (Kurpiewski et al., 1981; Rees et al., 1984; Tambourgi et al., 1998; Fernandes Pedrosa et al., 2002; Tambourgi et al., 2004). While the pathogenic bacterium

☆ GenBank data deposition information: all from *Loxosceles arizonica*: SMaseD cDNA 1, AF512953, SMaseD cDNA 2, AY699703, SMaseD cDNA 3, AY699704; genomic fragment 1 with 5′end and exon 1 of SMaseD homolog—AF512954; genomic fragment 2 with exons 5 and 6 (including 3′ end) of SMaseD homolog—AF512955; genomic fragments 3 including exons 3, 4, and 5 of SMaseD—AF512956.

* Corresponding author. Department of Biology, Lewis and Clark College, 0615 SW Palatine Hill Road, Portland, OR 97219, USA. Tel.: +1 503 768 7653; fax: +1 503 768 7658.

E-mail address: binford@lclark.edu (G.J. Binford).

0041-0101/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.toxicon.2004.11.011

Clostridium perfringens has been demonstrated to enhance the severity of lesions (Monteiro et al., 2002), it is clear from mammalian assays of SMaseD active cDNA expression products that SMaseD is the causative agent of lesion formation (Fernandes Pedrosa et al., 2002; Tambourgi et al., 2004). The mechanism by which cleavage of sphingomyelin (a ubiquitous eukaryotic membrane phospholipid) leads to severe tissue necrosis in humans is poorly understood but involves a complex immune response (Tambourgi et al., 1995, 1998, 2002; Desai et al., 1999). While cleavage of phospholipids is a common and necessary housekeeping phenomenon, cleavage at the D site (between the choline and phosphate) of these molecules is rare.

SMaseD activity is currently unknown in the animal kingdom outside of venoms in the *Loxosceles* lineage. Comparative analyses suggest a single evolutionary origin of SMaseD activity in the most recent common ancestor of *Loxosceles* and their sister genus *Sicarius* (Binford and Wells, 2003). Outside of this spider lineage, SMaseD activity is known in the pathogenic bacteria *Corynebacterium pseudotuberculosis*, *C. ulcerans*, *Archanobacterium haemolyticum* (formerly *Corynebacterium*) and *Vibrio damsela* (Bernheimer et al., 1985; Truett and King, 1993; Cuevas and Songer, 1993; McNamara et al., 1995). This disparate occurrence has led to speculation of an evolutionary relationship between the spider and bacterial SMaseDs (Bernheimer et al., 1985). SMaseDs from *Loxosceles*, *C. pseudotuberculosis*, and *A. haemolyticum* are similar in molecular weight (30–35 kDa), charge and isoelectric point, and share some conserved amino acid residues in the N-terminus (McNamara et al., 1995; Barbaro et al.,1996; Tambourgi et al.,1998; vanMeeteren, et al., 2004). SMaseD from *V. damsela* is larger than the other SMaseDs (69 kDa) and shares no other similarities outside of the SMaseD activity. Infection by *C. pseudotuberculosis* and envenomation by *Loxosceles* result in similar pathologies. Both have a comparable effect on neutrophils and the complement system (Yozwaik and Songer, 1993; Truett and King, 1993; Songer, 1997; Tambourgi et al., 2002; vanMeeteren, et al., 2004). However, *Loxosceles* and *Corynebacterium* SMaseDs are not antigenically cross-reactive (Bernheimer, et al., 1985).

There are three plausible evolutionary scenarios that could explain the similarities between spider and bacterial SMaseD: (a) bacterial and spider SMaseD could have independently evolved from the same general conserved protein family, (b) SMaseD could have originated in one lineage and moved to the other via horizontal transfer, (c) similarities between bacterial and spider SMaseD do not result from common ancestry but reflect convergence due to common function. Distinguishing among these mechanisms requires comparative analyses of SMaseD nucleotide and amino acid sequences, and characterization of SMaseD genes in *Loxosceles*.

Recently, cDNA homologs of SMaseD have been cloned and sequenced from two South American *Loxosceles* species (Fernandes Pedrosa et al., 2002; Kalapothakis et al., 2002; Tambourgi et al., 2004) (Fig. 1a). Here we report cDNA sequences from three SMaseD homologs and details about the genomic structure of SMaseD family members from the North American species *Loxosceles arizonica*. We subject these sequences to bioinformatic and phylogenetic analyses and discuss the implications on our understanding of the evolution of this unique venom toxin.

2. Methods

2.1. Spiders

Loxosceles arizonica, Arizona brown spiders (Gertsch and Ennik, 1983), were collected as adults from the desert of the Santa Catalina Mountain southern foothills (2500–2800’), Tucson, Arizona, Pima Co (32.1920N, 11.04833W) Voucher specimens are kept in the personal collection of GJB.

2.2. cDNA sequences

SMaseD homologous cDNAs were amplified using two degenerate primers that were designed from conserved

```
(a) L. intermedia* 1  ADKRRPIWNMGHMVNAIEQIDEFVNLGANNIEX-DVAF
    L. intermedia 1   AGNRRPIWIMGHMVNAIGQIDEFVNLGANSIEITDVFFD
    L. intermedia 1   ADKRRPIWIMGHMVNAIAQIDEFVNLGAN
    L. intermedia 2   AGNKRPIWIMGAMVNAIKDIXDFVNLGA-NNIX-K
    L. gaucho 2       ADNKRPIWVMGGMVNSLAQIKEFVGLGLDNSEKDNKWYK
    L. laeta 2        ADNRRPIWNKGHMVNALKQIPTFLXDGANA
    L. reclusa 3      AG-RRPVWIMGHMVNAIQQIDEFVNLGANAIEIDDVAWR
    L. reclusa 4      A-NKRPVWIMGHMVNAVYQIDEFVNLGANSIDTDVS
    L. deserta 4      A-NKRPIWIMGHMVNAIYQIDEFVNLGANSIDTDVS
    L. arizonica      A-NKRPIWIMGHMVNANYQIDEFVNLGANSIETDVSFDSS
              ruler 1.......10........20........30........40
```

Fig. 1. (a) Alignment of *Loxosceles* venom SMaseD N-terminal amino acid sequences (1, Tambourgi et al., 1998; 2, Barbaro et al., 1996; 3, Cisar et al., 1989; 4, Gomez et al., 2001). *L. intermedia** is a related venom protein without SMaseD activity (Tambourgi et al., 1998). Residues highlighted in grey are 100% conserved. (b). cDNA sequences and deduced amino acid sequences for three SMaseD paralogs from *Loxosceles arizonica.* The vertical line indicates the beginning of the mature protein. The bold and underlined 4 amino acid region NESA is a proposed glycosylation site.

(b)

```
L.arizonica 2 gttagagcaactgagaagttcgctcccatatacttttctgtcatccgttacagagtgcg    60
              V  R  A  T  E  K  F  A  P  I  Y  F  F  C  H  P  L  Q  S  A    20
L.arizonica 3 gttagagcaactgagaagttcgcttccatgtacttttctgtcattctccgcagagtgcc    60
              V  R  A  T  E  K  F  A  S  M  Y  F  F  C  H  S  P  Q  S  A    20

L.arizonica 2 gaaactgatgttgcagaacgc|gcaaataaacgacctatatggatcatgggacacatggtc   120
              E  T  D  V  A  E  R  |A  N  K  R  P  I  W  I  M  G  H  M  V   40
L.arizonica 3 gaaaccgatgttgcagaacgt|ggaaataaacgacctgtatggatcatgggacacatggtc   120
              E  T  D  V  A  E  R  |G  N  K  R  P  V  W  I  M  G  H  M  V   40

L.arizonica 2 aacgctaactatcagatagacgagtttgtgaaccttggagcgaattccattgaaacagac   180
              N  A  N  Y  Q  I  D  E  F  V  N  L  G  A  N  S  I  E  T  D   60
L.arizonica 3 aatgctatcgctcaaatagacgagtttgtgaaccttggagcgaattctattgaaacagac   180
              N  A  I  A  Q  I  D  E  F  V  N  L  G  A  N  S  I  E  T  D   60
L.arizonica 1           tcagattgacgagtttgtgaaccttggagcgaattccattgaaacagac    49
                          Q  I  D  E  F  V  N  L  G  A  N  S  I  E  T  D   16

L.arizonica 2 gtgtctttcgactccagtgccaatcctgaatatacgtatcacggcgttccatgcgactgt   240
              V  S  F  D  S  S  A  N  P  E  Y  T  Y  H  G  V  P  C  D  C   80
L.arizonica 3 gtctctttcgactccagtgccaatcctgaatacacgtatcacggcgttccatgcgattgt   240
              V  S  F  D  S  S  A  N  P  E  Y  T  Y  H  G  V  P  C  D  C   80
L.arizonica 1 gtgtctttcgactccagtgccaatcctgaatatacgtatcacggcgttccatgcgactgt   109
              V  S  F  D  S  S  A  N  P  E  Y  T  Y  H  G  V  P  C  D  C   36

L.arizonica 2 agaaggtggtgcaagaagtgggagtatttcaacaattttctaaaagctctgcgaaaagcc   300
              R  R  W  C  K  K  W  E  Y  F  N  N  F  L  K  A  L  R  K  A   100
L.arizonica 3 ggaaggacttgcacaaagtgggagcatttcaacgaatttctgaaaggtctgcgaaaagcc   300
              G  R  T  C  T  K  W  E  H  F  N  E  F  L  K  G  L  R  K  A   100
L.arizonica 1 agagggtggtgcaagaagtgggagtatttcaacaattttctaaaagctctgcgaaaagcc   169
              R  G  W  C  K  K  W  E  Y  F  N  N  F  L  K  A  L  R  K  A   56

L.arizonica 2 acaacaccaggtgactccaagtatcatgaaaagttagtgttagttgtatttgacctgaaa   360
              T  T  P  G  D  S  K  Y  H  E  K  L  V  L  V  V  F  D  L  K   120
L.arizonica 3 acgacaccaggcgactccaagtatcatgaaaagttagtgttagttgtatttgacctgaaa   360
              T  T  P  G  D  S  K  Y  H  E  K  L  V  L  V  V  F  D  L  K   120
L.arizonica 1 acaacaccaggtgactccaagtatcatgaaaagttagtgttagttgtatttgacctgaaa   229
              T  T  P  G  D  S  K  Y  H  E  K  L  V  L  V  V  F  D  L  K   76

L.arizonica 2 accggtagcctctacgataatcaagcttacgacgccggaaagaaattagcgaaaaatctc   420
              T  G  S  L  Y  D  N  Q  A  Y  D  A  G  K  K  L  A  K  N  L   140
L.arizonica 3 actggtagactctacgacaaccaagcttctgacgccggaaagaaattagcgaaaagtctc   420
              T  G  R  L  Y  D  N  Q  A  S  D  A  G  K  K  L  A  K  S  L   180
L.arizonica 1 accggtagcctctacgataatcaagcttacgacgccggaaagaaattagcgaaaaatctc   289
              T  G  S  L  Y  D  N  Q  A  Y  D  A  G  K  K  L  A  K  N  L   96

L.arizonica 2 cttcagcattactggaacaacggtaataatggggggaagagcatacatcgtattatccata   480
              L  Q  H  Y  W  N  N  G  N  N  G  G  R  A  Y  I  V  L  S  I   160
L.arizonica 3 cttcagaattactggaacaacggcaataatggtggaagagcatacatcgtattatccata   480
              L  Q  N  Y  W  N  N  G  N  N  G  G  R  A  Y  I  V  L  S  I   200
L.arizonica 1 cttcagcattactggaacaacggtaataatggggggaagagcatacatcgtattatccata   349
              L  Q  H  Y  W  N  N  G  N  N  G  G  R  A  Y  I  V  L  S  I   116

L.arizonica 2 ccaaaccttgctcattataaattaattactggatttaaagaaacgctgaagaccgagggg   540
              P  N  L  A  H  Y  K  L  I  T  G  F  K  E  T  L  K  T  E  G   180
L.arizonica 3 ccaaaccttgcccattataaattaattgctggatttaaagaagcgcttacaagcgagggg   540
              P  N  L  A  H  Y  K  L  I  A  G  F  K  E  A  L  T  S  E  G   220
L.arizonica 1 ccaaaccttgctcattataaattaattactggatttaaagaaacgctgaagaccgagggg   409
              P  N  L  A  H  Y  K  L  I  T  G  F  K  E  T  L  K  T  E  G   136

L.arizonica 2 catccagagttgatggagaaagttggttatgacttttctggaaacgataacatcgaccaa   600
              H  P  E  L  M  E  K  V  G  Y  D  F  S  G  N  D  N  I  D  Q   200
L.arizonica 3 catccagaattgatggacaaagttggttatgacttttctggaaacgatgacatcggcgac   600
              H  P  E  L  M  D  K  V  G  Y  D  F  S  G  N  D  D  I  G  D   200
L.arizonica 1 catccagagttgatggagaaagttggttatgacttttctggaaacgataacatcgaccaa   469
              H  P  E  L  M  E  K  V  G  Y  D  F  S  G  N  D  N  I  D  Q   156
```

Fig. 1 (*continued*)

```
L.arizonica 2 gtcgcgaatgcgtacaagaaagctggagtgaccgggcatgtgtggcagagcgatggcatt   660
              V  A  N  A  Y  K  K  A  G  V  T  G  H  V  W  Q  S  D  G  I     220
L.arizonica 3 gtcgcgaatgcttacaagaaagccggagtaacagggcatgtgtggcagagcgatggcatc   660
              V  A  N  A  Y  K  K  A  G  V  T  G  H  V  W  Q  S  D  G  I     220
L.arizonica 1 gtcgcgaatgcttacaagaaagccggagtaacagggcatgtgtggcagagcgatggcatc   529
              V  A  N  A  Y  K  K  A  G  V  T  G  H  V  W  Q  S  D  G  I     176

L.arizonica 2 acgaactgtgtagcctcatttattcgcggacttgatcgcgcgaagaaagctgtgaaaaac   720
              T  N  C  V  A  S  F  I  R  G  L  D  R  A  K  K  A  V  K  N     240
L.arizonica 3 acaaactgt---------ttactgcggggtcttgatcgtgtggggaaagctgttgcaaac   711
              T  N  C  -  -  -  L  L  R  G  L  D  R  V  G  K  A  V  A  N     237
L.arizonica 1 acaaactgt---------ttactgcggggtcttgatcgtgtgaggaaagctgttgcaaac   580
              T  N  C  -  -  -  L  L  R  G  L  D  R  V  R  K  A  V  A  N     193

L.arizonica 2 agagattcttcaaacggatacattaacaaagtgtactattggacagtggacaagtacgca   780
              R  D  S  S  N  G  Y  I  N  K  V  Y  Y  W  T  V  D  K  Y  A     260
L.arizonica 3 agagattcttcaaacggatacattaacaaagtgtactattggacagtggacaagcgccaa   771
              R  D  S  S  N  G  Y  I  N  K  V  Y  Y  W  T  V  D  K  R  Q     257
L.arizonica 1 agagattcttcaaacggatacattaacaaagtgtactattggacagtggacagcgccaaa   640
              R  D  S  S  N  G  Y  I  N  K  V  Y  Y  W  T  V  D  K  R  Q     213

L.arizonica 2 acgactagagaagcattcgacattggagtcgatggaataatgaccaattacccggatgtc   840
              T  T  R  E  A  F  D  I  G  V  D  G  I  M  T  N  Y  P  D  V     280
L.arizonica 3 tcgactagagatgcactcgatgctggagtcgatggaataatgaccaattacccggatgtt   831
              S  T  R  D  A  L  D  A  G  V  D  G  I  M  T  N  Y  P  D  V     277
L.arizonica 1 tcgactaaaaatgcactcgatgctggagtcgatggaataatgcccaattacccggatgtt   700
              S  T  K  N  A  L  D  A  G  V  D  G  I  M  P  N  Y  P  D  V     233

L.arizonica 2 attgctaatgtcctcaatgaatctgcttataaagggaaattcagacttgccacatacgac   900
              I  A  N  V  L  N  E  S  A  Y  K  G  K  F  R  L  A  T  Y  D     300
L.arizonica 3 attgctgatgtcctcaatgaatctgcttataaagcgaaattcaggattgcctcatacgac   891
              I  A  D  V  L  N  E  S  A  Y  K  A  K  F  R  I  A  S  Y  D     297
L.arizonica 1 attgctgatgtccccaatgaatctgcttataaagcgaaattcaggattgcctcatacgac   760
              I  A  D  V  P  N  E  S  A  Y  K  A  K  F  R  I  A  S  Y  D     253

L.arizonica 2 gacaatccttgggaaacattcaagaatta                                  929
              D  N  P  W  E  T  F  K  N  -                                   309
L.arizonica 3 gacaatccttgggaaacatacaagaatta                                  920
              D  N  P  W  E  T  Y  K  N  -                                   306
L.arizonica 1 gacaatccttgggaaacattcaagaatctgtaggtttactctgcggatcgcattgcaatc   820
              D  N  P  W  E  T  F  K  N  -                                   262

L.arizonica 1 cggatttctccttgaactttaaaaaaacgtttaatgtgcatgaaaaaaaataattcagtt   880
L.arizonica 1 ataatatgtgaaaaaggtctttagcaatttgtaaatcgatcttcttcaataaaatatttg   940
L.arizonica 1 atagctt                                                        947
```

Fig. 1 (*continued*)

regions of N-terminal amino acid sequences of purified proteins with known SMaseD activity (Fig. 1a). Venom glands were dissected out of 40 *L. arizonica* two days after venom was removed from glands by electrostimulation, corresponding to a time when we expected SMaseD expression. mRNA was isolated from the venom gland tissue using a Clontech Nucleic Acid Purification kit. cDNA was generated by RT-PCR using Superscript II (Gibco), and SMaseD related cDNAs were amplified using the upstream degenerate primer (5′-TGGATHATGGGNCAYATGGT-3′) and a polyT primer with a random 20mer. Products of this reaction were amplified with the nested degenerate primer (5′-CARATHGAYGARTTYGT-3′), cloned (Invitrogen TA cloning) and sequenced using standard techniques.

Sequence upstream of the primers was identified using genomic library screens (see Section 2.3). Primers for amplifying the complete coding region (signal peptide and mature protein) were designed from the genomic sequence for the N-terminus (5′-GTTTCCATGGTTAGAGCAACT-GAGA-3′, underlined sequence is a NcoI site with a 5 base pair adapter sequence), and from the cDNA sequence for the 3′ antisense C-terminus (5′ TTTTCTCGAGT-TAATTCTTGAATGTTTCCCA-3′, underlined sequence is an XhoI site with a four base pair adapter sequence). PCR products resulting from amplification of venom gland cDNA using these primers were submitted for direct sequencing and cloned (Invitrogen TA vectors). Cloned fragments were sequenced using M13F and M13R primers.

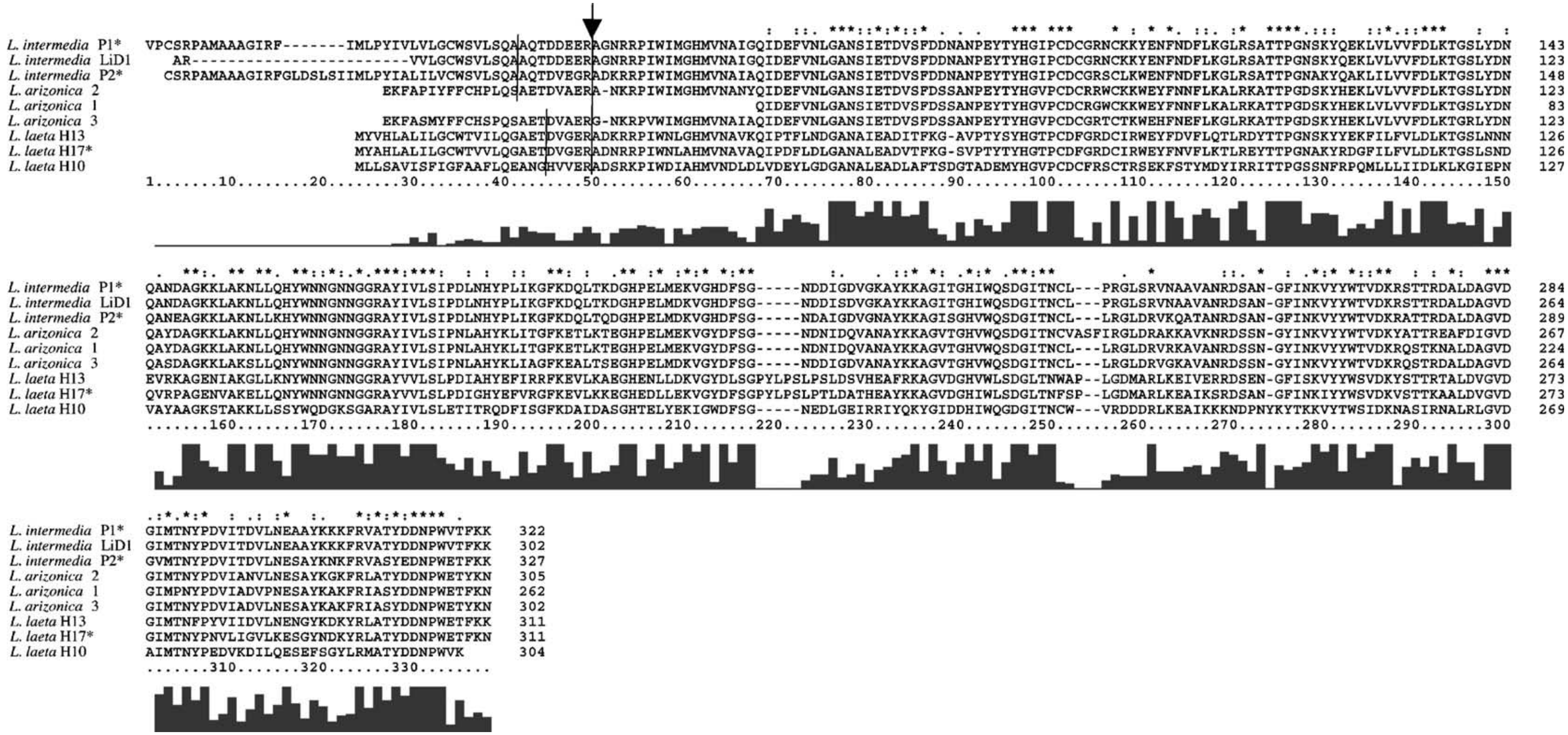


Fig. 2. Alignment of amino acid sequences for all available SMaseDs. The histogram reflects degree of conservation. The first vertical lines for each sequence indicate estimated cleavage sites between signal and propeptides (SignalP 3.0, DyrlØv Bendtsen et al., 2004). The second vertical line with the arrow above indicates the beginning of the mature protein. GenBank accession numbers are as in Table 2. Asterisks indicate names of sequences whose expression products have been demonstrated to exhibit SMaseD activity.

2.3. Gene structure

To characterize the gene structure of SMaseD we made and screened a phage genomic library with SMaseD cDNA labeled with digoxigenin (DIG) The phage genomic DNA library was created using ZAP Express (Stratagene) from 450 mg of *L. arizonica* tissue. DNA was digested overnight in 6 U/μl Sau3A1 then size selected for an average insert size of 5 kb using sucrose gradient centrifugation. A SMaseD probe was made by incorporating DIG into the cDNA sequence using random-primed labeling (DIG High Prime, Boehringer Mannheim Corp.) (Roche). Plaques were grown on NZY plates ((50,000 pfu/plate), lifted onto Magna, Nylon transfer membranes (0.45 micron, Osmonics, Inc.), UV crosslinked and incubated overnight with DIG-labeled SMaseD cDNA. Membranes were washed and incubated with DIG antibody. Positive clones were excised, purified and cloned into pBluescript II KS(+/-).

Three library inserts (~5 kb) with homology to the cDNA were determined to be distinct from one another by restriction digest analyses (EcoR1, Kpn1, Xba1). These three genomic fragments were sequenced using primer walking and assembled by identifying overlapping regions. Exons were identified by pairwise comparisons of the genomic DNA with the cDNA. Attempts to use PCR to amplify the region including exon 2 were unsuccessful.

2.4. Sequence analyses

cDNA, amino acid, and genomic sequences were analyzed for homology to known sequences using NCBI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST). Protein domain family analysis was done using Pfam (http://pfam.wustl.edu). Fold recognition analysis was performed by submitting Clustal-generated multiple sequence alignments to the 3D-PSSM server (http://www.sbg.bio.ic.ac.uk/ servers/3dpssm/) (Fischer et al., 1999; Kelly et al., 1999). Pairwise global alignments were performed with GeneStream Align (http://xylian.igh.cnrs.fr/). N-glycosylation sites were identified using PROSITE NiceSite (http://ca.expasy.org/cgi-bin/nicedoc.pl?PDOC00001). Signal peptide cleavage sites were estimated using SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/)(DyrlØv Bendtsen et al., 2004).

2.5. Phylogenetic analysis

Evolutionary relationships among amino acid SMaseD cDNA sequences, and exon 5 (nucleotide only) from *L. arizonica*, *L. laeta*, and *L. intermedia* (see Fig. 2) were analyzed using Bayesian (MrBayes, Huelsenbeck and Ronquist, 2001), parsimony and neighbor-joining (PAUP*, Swofford, 2001) phylogenetic algorithms. Nucleotide sequences were also analyzed using maximum likelihood (PAUP*, Swofford, 2001). All sequences were aligned using ClustalX with gap and extension penalties of 10 and 0.1, respectively, and were corrected by hand using MacClade 4.02 (Maddison and Maddison, 2001). Amino acid alignments were weighted using the Gonnet series scoring matrix. Analyses with other gap penalty setting and weighting schemes resulted in the same tree topology. Maximum likelihood used GTR and gamma scale parameters estimated on a single most parsimonious tree. (Parameter values: AC = 1.5, AG = 2.8, AT = 1.6, CG = 1.7, CT = 3.9, GT = 1.0; $A = 0.32\%$, $C = 0.20\%$; $G = 0.23\%$; $T = 0.26\%$; $\alpha = 1.19$.) Robustness of parsimony and maximum likelihood tree topologies was assessed with bootstrap analyses (1000 and 100 replicates, respectively) and posterior probabilities (Bayesian, 10,000,000 generations, saved every 100th tree, 1000 burnin).

3. Results and discussion

3.1. cDNA sequences

Our analyses yielded two distinct cDNAs that include a partial signal peptide, a propeptide and the complete mature protein (cDNA 2 = AY699703, cDNA 3 = AY699704) and one partial cDNA that does not include the N-terminus of the mature peptide, but does include the 3′ untranslated region (cDNA 1 = AF512953) (Fig. 1b). Sequence homology and estimated molecular weights substantiate the identity of these sequences as mRNAs of members of the SMaseD family. Deduced amino acid sequences and alignments to SMaseD protein sequences from other *Loxosceles* species are shown in Fig. 2. The three *L. arizonica* cDNA sequences are sufficiently different that they likely represent three paralogous members of a gene family (Table 1). BLASTn searches of the cDNA sequences identified homology only to SMaseD cDNAs recently determined from South American *Loxosceles* species (Fig. 2). The estimated molecular weights of the deduced protein sequence are comparable to mature SMaseDs purified from *Loxosceles* venoms (cDNA 2 = 32,086 Da, cDNA 3 = 31,041 Da) (Kurpiewski et al., 1981; Rees et al., 1988; Tambourgi et al., 1998; Fernandes Pedrosa et al., 2002; Tambourgi et al., 2004). Outside of *Loxosceles* there were no homologous sequences or fragments larger than 21 nucleotides.

BLASTp searches of the deduced amino acid sequence show strong similarity (45–92%; *E* values from 0.08 to

Table 1
Percent nucleotide identity between SMaseD cDNA sequences from *Loxosceles arizonica*. In parentheses is the number of overlapping nucleotides in the pairwise comparison

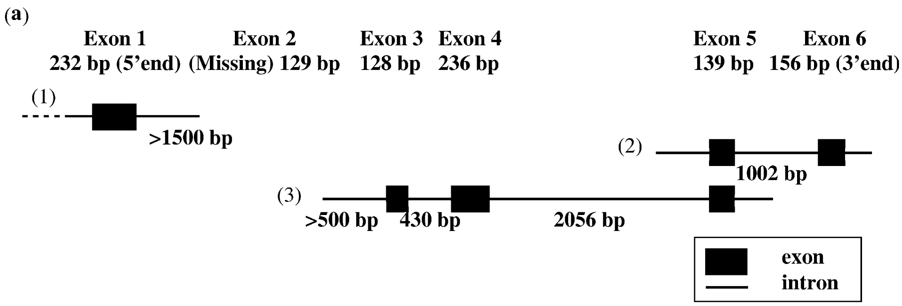
%Nucleotide identity	cDNA 2	cDNA 3
cDNA 1	94.9 (800)	93.7 (791)
cDNA 2		90.2 (949)

Table 2
Pairwise global percent amino acid identity between protein sequences for: *L. arizonica* SMaseD paralogs (cDNA 1, AF512953, cDNA 2, AY699703, cDNA 3, AY699704), *L. intermedia* (*L.int*) SMaseD paralogs (LiD1, AY340702, P1, AY304471, P2, AY304471), *L. laeta* SMaseD paralogs (H17, AY093599, H13, AY093600, H10, AY093601), *Corynebacterium ulcerans* (*C. ulc.*) SMaseD (Q59332), *Corynebacterium glutamicum* (*C. glut.*) GDPD (NP_602097), *Drosophila melanogaster* (*D. mel.*) GDPD (AAF55318), human GDPD (AAL55858)

%Amino acid identity	cDNA 2	cDNA 3	*L. int.* LiD1	*L. int* P1	*L. int* P2	*L. laeta* H17	*L. laeta* H13	*L. laeta* H10	*C. ulc.* SMaseD	*C. glut.* GDPD	*D. mel* GDPD	Human GDPD
cDNA 1	94.0	90.5	81.2	80.9	79.8	59.0	58.2	44.2	21.6	18.6	18.6	17.4
cDNA 2		89.4	73.8	72.0	68.8	58.8	58.4	43.6	21.9	19.5	14.3	15.4
cDNA 3			77.1	74.5	72.8	55.3	56.3	43.1	24.0	18.9	17.9	16.5
LiD1				92.5	81.3	59.1	58.7	43.0	20.7	19.9	12.3	16.8
L. int P1					86.9	56.1	55.8	41.2	20.7	20.3	17.7	17.3
L. int P2						55.0	55.6	39.9	19.5	19.2	17.5	16.2
L. laeta H17							80.9	42.4	23.8	19.7	17.6	19.1
L. laeta H13								40.9	20.9	19.4	17.6	18.7
L. laeta H10									22.8	17.6	17.6	17.3
C. ulc SMaseD										17.0	14.5	15.7
C. glut GDPD											18.2	17.8
D. mel GDPD												21.3

3e-11) only with partial and complete amino acid sequences from *Loxosceles* (Figs. 1a and 2). Also identified by BLASTp, however, were several weak hits ($E=0.1$–5.7) of the SMaseD C-terminal region (residues 210–250) to the C-termini of a variety of eukaryotic and prokaryotic sequences that have been annotated as glycerophosphoryl diester phosphodiesterases (GDPD pfam03009). The similarity between the reactions catalyzed by these different enzymes suggested to us that the weak BLAST hits might be the result of a distant evolutionary relationship between SMaseD and the GDPDs. In support of this hypothesis, a Pfam HMM search with *Loxosceles* SMaseD revealed statistically significant local similarity ($E=0.00011$) to the GDPD domain profile in the C-terminal region and weaker similarity to the global profile ($E=9.0$). SMaseDs (282 and 279 residues) were also similar in overall size to the GDPD domain (average length 235 residues). Moreover, submission of a multiple alignment of known SMaseDs (Fig. 2) to the fold recognition program 3D-PSSM yielded the only known structure of a GDPD family member (1olz; protein from the bacterium *Thermotoga maritima*) as the single best match with an E-value$=0.231$. The *Thermotoga maritima* protein has a TIM barrel structure, a fold commonly found for enzymes. Together, the above findings strongly support that *Loxosceles* SMaseD is a single domain TIM barrel protein and a derived member of the broadly conserved GDPD protein domain family.

Earlier studies have noted similarity between N-terminal amino acid sequences from *Loxosceles* SMaseD and the N-termini of a group of toxic bacterial phospholipase D (SMaseD) enzymes (McNamara et al., 1995; Tambourgi et al., 1998). Comparison of full-length spider and bacterial SMaseD sequences here and in other recent work (Van Meeteren et al., 2004) provide no evidence to support the conclusion that these proteins are related to each other. Bacterial SMaseDs do not show up as significant hits in BLAST searches with full length *Loxosceles* SMaseD. Global pairwise alignments between *Loxosceles* SMaseD and bacterial SMaseDs yield sequence identities of 21–24% (Table 2) and alignment scores ranging from 95–132, values that do not represent significant evidence for homology. Moreover, although *Loxosceles* SMaseD is shown above to be a diverged member of the GDPD family, bacterial SMaseDs are not recognized as GDPD family members by Pfam searches, or by fold recognition analysis using the 3D-PSSM server. Thus, the bacterial and spider proteins cannot be shown to be homologous indirectly through a shared relationship to the GDPD family. However, there is also no evidence to rule out the possibility that *Loxosceles* SMaseD and the bacterial SMaseDs are both derived GDPD family enzymes and share a more recent common ancestor with each other than either does with the broadly conserved members of this family. Although SMaseD shows significant similarity to the GDPD domain profile, its divergence from this group of proteins is extreme. Overall sequence identities from global alignments with various eukaryotic



(b)

Genomic fragment	1	2			3	
cDNA sequence	Exon 1 (309)	Exon 3 (128)	Exon 4 (236)	Exon5 (143)	Exon5 (143)	Exon 6 (156)
1	98.9 (180)	94.5	89.0	91.6	97.9	97.4
2	100 (309)	94.5	93.2	92.6 (129)	94.6 (129)	
3	89.2 (308)	90.6	87.7	91.7 (132)	98.5 (132)	

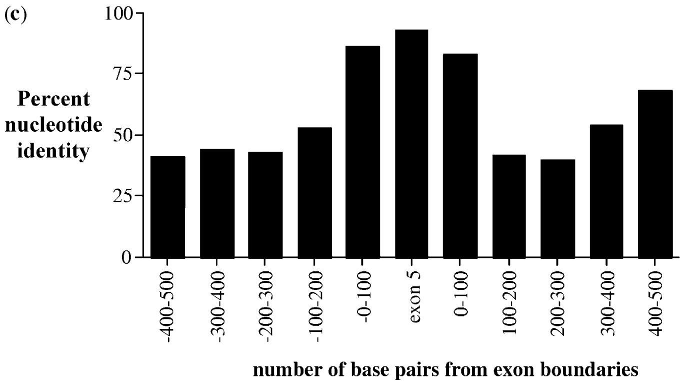


Fig. 3. (a) Genomic structure of exons with homology to SMaseD cDNA. Individual genomic fragments are numbered (1–3). (b). Summary of sequence identity between cDNA and segments of genomic DNA inserts with homology to the cDNA. Numbers in parentheses correspond to the length of overlap between the exons and available cDNAs. (c). Percent nucleotide identity of exon 5 and flanking intron regions of increasing distance up and downstream of the exon between two independent genomic library inserts (2 and 3) on which this exon was present.

members of this family are quite low (14–21%, Table 1b) and alignment scores are also quite poor (−111 to 50). Sequence identities and global alignment scores between SMaseD and bacterial SMaseDs, while also low, are actually slightly higher than this, leaving open the possibility that the common function of these enzymes derives from the same ancient divergence from the GDPD family.

3.2. Gene structure

Genomic library screens identified three genomic fragments (Fragment 1 *AF512954* 4652 bp, 2 *AF512955* 4080 bp, and 3 *AF512956* 4686 bp) containing portions of SMaseD homologous genes (Fig 3). Pairwise comparisons of cDNA sequences and the genomic fragments identified 5 exons with a sixth region of the cDNA that did not match any of the genomic DNA ('exon 2'). Attempts to PCR amplify 'exon 2' from genomic DNA using primers generated from exons 1 and 3 were unsuccessful as were attempts to amplify using primer pairs from exons 1 and 2, and exons 2 and 3. Similarity between the cDNA sequences and exons (genomic DNA) is summarized in Fig. 3b. Genomic fragments 2 and 3 share a common exon (exon 5, 94% sequence identity) but introns are increasingly divergent with distance from the exon (Fig. 3c) and thus represent two paralogs from the SMaseD gene family (Fig. 3a). Using overlap between these paralogs (exon 5 and surrounding sequence) and assuming conservation of intron/exon structure within the family, we estimate that members of this gene family minimally span 6500 base pairs and contain at least 6 exons and five introns (Fig. 3a).

Genomic fragment 1 includes the N-terminus and upstream sequence of the SMaseD gene. Putative regulatory elements (Fig. 4) include a capsite that plays a role in initiation of transcription and is found in many arthropod genes (Cherbas and Cherbas, 1993). The 3′ untranslated region includes the polyadenylation signal, AATAAA,

```
AAACAAACAAAAACTCAGTCAAGAATGTCCGAAAACACTTTGAATCTGCTTATTACGTGCAGG

GTGAATGTGTTTTCATGTAACCCAATACGATTTGACTAAAGAAAACGGTTTTGTAGGAGAGAA

AGGTTATATATATAATTCAGCTTGATAAAATGTCTCATAGTTCTACTGCTCTACTTCATCCAT
                                 M~~S~~H~~S~~S~~T~~A~~L~~L~~H~~P~~Y

ACGTAGCTGCAAGAGCAACTGAGAAGTTCGCTCCCATATACTTTTTCTGTCATCCGTTACAGA
~~V~~A~~A~~R~~A~~T~~E~~K~~F~~A~~P~~I~~Y~~F~~F~~C~~H~~P~~L~~Q~~S

GT|GCGGAAACTGATGTTGCAGAACGCGCAAATAAACGACCTATATGGATCATGGGACACATGG
~~|A~~E~~T~~D~~V~~A~~E~~R~~A~~N~~K~~R~~P~~I~~W~~I~~M~~G~~H~~M~~V

TCAACGCTAACTATCAGATAGACGAGTTTGTGAACCTTGGAGCGAATTCCATTGAAACAGACG
~~N~~A~~N~~Y~~Q~~I~~D~~E~~F~~V~~N~~L~~G~~A~~N~~S~~I~~E~~T~~D~~V
```

Fig. 4. Upstream sequence of 5′ end of SMaseD from genomic DNA. Underlines indicate, in order: putative capsite consensus (Cherbas and Cherbas, 1993), CAAT box, TATATA box, initiating methionine, and the beginning of exon 1. The arrow indicates the putative cleavage site for the signal peptide (DyrlØv Bendtsen et al., 2004).

13 bp upstream of the polyA tail. The 5′ region upstream of the mature protein shows a signal peptide and a recognizable propeptide sequence consistent with expression of SMaseD as a zymogen, which is typical of eukaryotic digestive enzymes. Zymogens are expressed as inactive precursors and require trypsin cleavage of a signal peptide for activation. This may provide a mechanism of control of the timing and circumstance of activation of potentially tissue-destructive activity. Analyses of all SMaseD cDNAs with the SignalP 3.0 signal peptide recognition algorithm (DyrlØv Bendtsen et al., 2004) estimate two different cleavage sites between the signal and the propeptides (Fig. 2) resulting in 5 or 8 amino acid propeptides. Kalapothakis et al. (2002) estimated a cleavage site for *L. intermedia* LiD1 that results in a 10 amino acid propeptide. The difference is likely due to the use of different algorighms for estimation. All proposed propeptides are consistent with sizes of propeptides found in other spider venom toxins (Santos et al., 1992; Diniz et al., 1993; Kalapothakis et al., 1998, 2002).

In congruence with reports that glycosylation is necessary for activity of *Loxosceles* venom components (Viega et al., 1999) the *L. arizonica* cDNAs each have one potential glycosylation site (NESA) that is conserved among paralogs in this species and shared with confirmed SMaseD active *L. intermedia* P2 (Figs. 1b and 2). Residues 257–260 of *L laeta* 3 (NASI) also constitute a potential glycosylation site. Interestingly, *L. laeta* 1 and *L. intermedia* P1, both with confirmed SMaseD activity (Fernandes Pedrosa et al., 2002; Tambourgi et al., 2004, respectively), do not have potential N-glycosylation sites.

3.3. Molecular evolution

3.3.1. Diversification of the SMaseD gene family

The presence of three paralogs of SMaseD in *L. arizonica* venom gland cDNA adds to mounting evidence that multiple members of this gene family are expressed in *Loxosceles* venoms. In fact, the number of expressed paralogs in venoms is independently converging on three by different research groups. Three paralogous sequences were recently determined from *L. laeta* (Fernandes Pedrosa et al., 2002) and *L. intermedia* venom gland cDNA (Kalapothakis et al., 2002; Tambourgi et al., 2004). Furthermore, separation of proteins in crude *Loxosceles* venoms have identified multiple active forms of SMaseD within venoms of single *Loxosceles* species and expressed proteins with homology to SMaseD but without SMaseD activity (Kurpiewski et al., 1981; Tambourgi et al., 1998; Fig. 1a). It is likely that both expression of multiple paralogous SMaseD-related genes and post-translational modification contribute to differences in composition of venoms within this lineage.

One hundred percent nucleotide sequence identity between exon 1 on genomic fragment 1 and cDNA 2 is strong evidence that fragment 1 contains the gene coding for the mRNA for cDNA 2. No other exons are perfect matches between genomic and cDNA sequences. This raises the possibility that the *L. arizonica* SMaseD gene family contains more paralogs than the three we isolated from venom gland expression products. In fact, genomic fragment 2 has percent nucleotide identities with cDNAs as low as 87.7% and reaching only as high as 94.5% (Fig. 3b). Unless they are undergoing some rare circumstance like RNA editing, these new sequences suggest that the gene family contains at least 4 or 5 members. It also suggests that some of these members are not expressed in the venom gland. While we cannot rule out the possibility that other paralogs are expressed in the venom glands, we have some confidence that there are only three because we have done sufficient RT-PCR from venom gland mRNA to obtain the same paralogs multiple times.

Phylogenetic analyses of all available SMaseD cDNAs provide strong support that paralogs within all *Loxosceles* species included in this analysis are more closely related to other paralogs in the same species than they are to homologous sequences from any other species (Fig. 5). This could be explained by either duplication events that

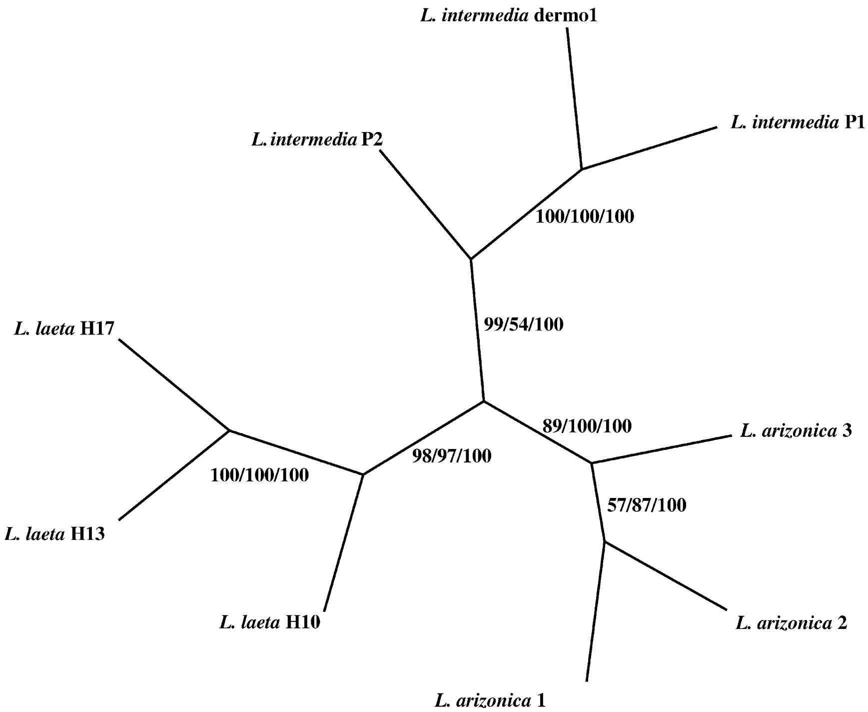


Fig. 5. Unrooted tree topology resulting from all analyses of nucleotide and amino acid sequences of SMaseD. Numbers on branches are parsimony bootstrap values (amino acid) (1000 reps)/maximum likelihood bootstrap values (nucleotide) (1000 reps)/Bayesian posterior probabilities (amino acid) (10,000,000 generations and 4 chains).

have occurred independently since the most recent common ancestors of each taxon pair, or by concerted evolution homogenizing the paralogs within species. Since researchers have independently found three paralogs in all species studied, concerted evolution is the more likely mechanism as it is unlikely that duplication events would have independently converged on three copies in each lineage. Concerted evolution would result in patterns of molecular similarity in which antibodies raised against particular species would likely be most effective against closely related species. Furthermore, it would increase the probability that a monoclonal antibody would be effective against all paralogs within an individuals' venom.

4. Mechanism of evolutionary origin of Loxosceles SMaseD

High levels of divergence between SMaseD and all other sequences in GenBank, and the lack of sequence information from other chelicerate taxa, make deciphering the evolutionary origins of SMaseD in *Loxosceles* challenging. However, characteristics of the gene structure (introns, signal elements), and evidence that SMaseD is a member of a gene family make it clear that this unique toxic enzyme has been evolving within a eukaryotic genome for a long time. This rules out the hypothesis of SMaseD having originated in *Loxosceles* via *recent* horizontal transfer of bacterial SMaseD*s*. Further, remote homology between *Loxosceles* SMaseD and the ubiquitous GDPD family suggests that it could have arisen by duplication and divergence of a housekeeping gene with similar cleavage activity.

However, the sequence data presented here neither support nor rule out origin of SMaseD by an *ancient* horizontal transfer event. *Loxosceles* SMaseD shows only marginally greater global sequence identity with bacterial SMaseD exotoxins than with eukaryotic GDPD family members (Table 2). These levels of identity lie in a region of similarity in which identification of homology among amino acid sequences is tenuous (Brenner et al., 1998). The only strongly supported conclusion of evolutionary significance from the cDNA sequence comparisons is that spider SMaseDs are divergent GDPDs. Bacterial SMaseDs show no significant relationships with anything in current databases. The slightly higher similarity between bacterial and spider SMaseDs than either of these shares with any GDPD (Table 1b) could be explained from limited sequence convergence due to evolution of a similar biological function. Thus, the present data cannot unequivocally distinguish between competing hypotheses of a single evolutionary origin of SMaseD activity followed by ancient horizontal transfer, or independent origins of SMaseD activity and a small amount of sequence convergence due to functional constraints. Phylogenetic analyses including spider and bacterial SMaseDs and GDPDs that would ideally help to identify occurrences of horizontal gene transfer are not possible because the degree of divergence between these sequences make alignments evolutionarily meaningless.

While the fingerprint of an ancient horizontal transfer of an SMaseD gene may be rendered difficult to uncover by the long evolutionary time span and sparse taxon coverage of related genes in current databases, there are some general aspects of the spider and bacterial lineages that are relevant to the issue. Comparative enzyme analyses support a single origin of SMaseD activity in venoms in the most recent common ancestor of *Loxosceles* and their sister genus *Sicarius* (Binford and Wells, 2003). The age of this lineage is unknown but their biogeographic distribution suggests it could be as old as 150–250 million years. An ancient event of horizontal transfer would mean the ancestral SMaseD of bacterial origin would have had to acquire its current eukaryotic properties (accumulated introns, acquired classic eukaryotic regulatory elements, and undergone duplication events) in this amount of time. Mounting evidence indicates that rates of evolutionary gains and losses of introns vary greatly among lineages of genes with some being quite high (Robertson, 1998; Tarrio et al., 1998; Boudet et al., 2001; Kiontke et al., 2004). Thus, it is not out of the question that introns could have been acquired in the suspected time frame after lateral transfer from bacteria.

General aspects of *Corynebacteria* are relevant to understanding the likelihood of a horizontal transfer event involving these bacteria. *Corynebacteria* are free-living and easily isolated from the soil and, thus, an ancient association between the ancestral *Loxosceles*/*Sicarius* and ancestral *Corynebacteria* is not unrealistic. Furthermore, it is clear that horizontal gene transfer is rampant among prokaryotes (see, for example, Gogarten

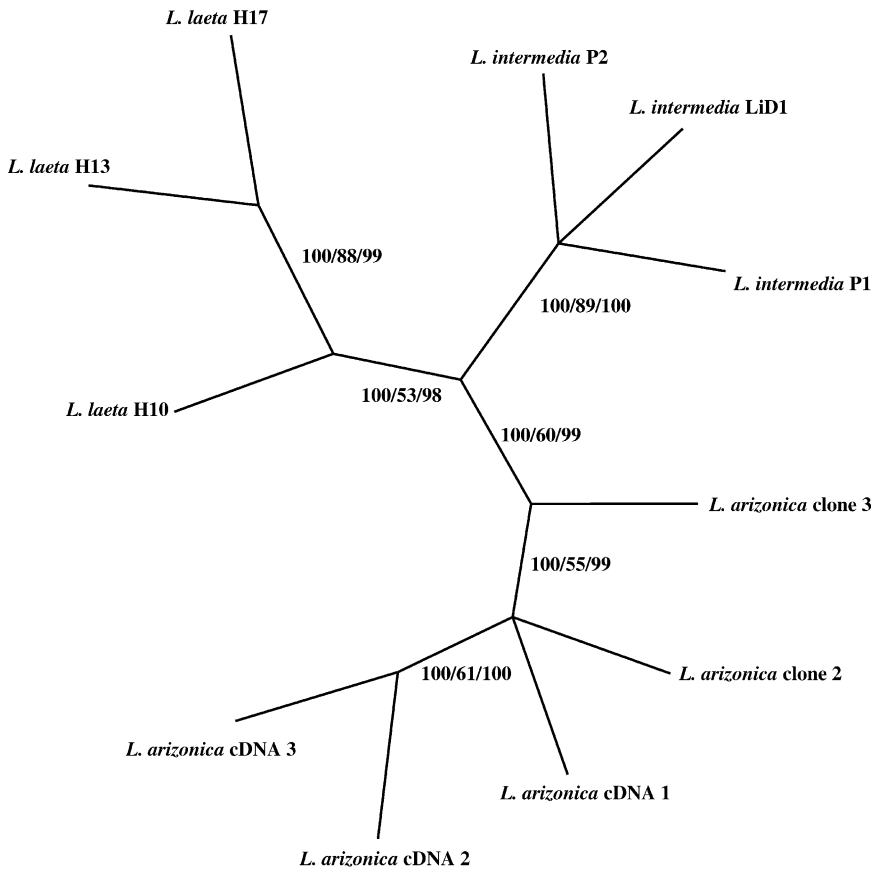


Fig. 6. Unrooted tree topology that resulted from all analyses of nucleotide sequences of exon 5. Numbers on branches are parsimony bootstrap values (1000 reps)/maximum likelihood bootstrap values (1000 reps)/ Bayesian posterior probabilities (10,000,000 generations and 4 chains).

et al., 2002) with pathogenicity-related genes being more readily transferred than others (Nakamura, 2004). The *Corynebacterium* lineage, however, has atypically high conservation of gene order, perhaps resulting from lack of recombinatorial repair genes (Nakamura et al., 2003) which would not likely be true of genomes with a propensity toward rampant horizontal gene transfer.

Ultimately, the data required to confidently reject the hypothesis of ancient horizontal gene transfer in favor of independent *de novo* origin of SMaseD activity lie in the distribution and properties of homologs of SMaseD in *Loxosceles* and close relatives without SMaseD activity. SMaseD has not been detected in venoms or tissues outside of close relatives of the *Loxosceles*/*Sicarius* lineage (Binford and Wells, 2003). Furthermore, preliminary genomic Southerns have not uncovered related sequences in the genomes of close relatives of *Loxosceles*/*Sicarius* (unpublished data). Unfortunately both of these pieces of evidence are enigmatic negative results. The identification of a SMaseD paralog expressed in *L. intermedia* venom that does not have SMaseD activity (Tambourgi et al., 1998) is provocative, but ultimately does not help clarify the issue without knowing if the present function of that paralog is ancestral or derived with respect to SMaseD. The same argument applies to *L. arizonica* genomic fragments that are SMaseD paralogs that may not be expressed in the venom glands. Phylogenetic analyses of exon 5 consistently reconstruct genomic fragment 3 to be ancestral to the three venom-expressed paralogs (Fig. 6). If this gene codes for a protein that is not a functional SMaseD, this would constitute strong evidence that this toxic activity emerged from within the genome of the ancestors of *Loxosceles*. Efforts are underway to isolate and obtain the complete cDNA sequence for this paralog.

A manuscript published while this work was in review by Ramos-Cerillo et al. (2004) (Toxicon 44:507–614) reports new sequences homologous to SMaseD from two North American species, *Loxosceles boneti* and *Loxosceles reclusa*. One of the paralogs from *Loxosceles boneti*, boneti 3, shows considerable divergence from all other SMaseD homologs from North American species and does not exhibit SMaseD activity. This and the divergent *L. laeta* sequence H10 suggest that venom-expressed members of this gene family that have a different enzyme activity are not being homogenized with their SMaseD active paralogs by concerted evolution.

Acknowledgements

We thank James Pennington and Jun Isoe for help in the lab and Margaret Kidwell and two anonymous reviewers for comments on the manuscript. This work was supported by a NIH training grant to Center for Insect Sciences, University of Arizona (L. Tolbert P.I.), NIH NRSA fellowship F32 GM20874-01 to GJB and MAW, and start-up funds from Lewis and Clark College.

References

Barbaro, K.C., Sousa, M.V., Morhy, L., Eickstedt, V.R.D., Mota, I., 1996. Compared chemical properties of dermonecrotic and lethal toxins from spiders of the genus *Loxosceles* (Araneae). J. Protein Chem. 15, 337–343.

Bernheimer, A.W., Campbell, B.J., Forrester, L.J., 1985. Comparative toxinology of *Loxosceles reclusa* and *Corynebacterium pseudotuberculosis*. Science 228, 590–591.

Binford, G.J., Wells, M.A., 2003. The phylogenetic distribution of sphingomyelinase D activity in venoms of Haplogyne spiders. Comp. Biochem. Physiol. B 135, 25–33.

Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M., Lecharny, A., 2001. Evolution of intron/exon structure of DEAD helicase family genes in *Arabidopsis Caenorhabditis*, and *Drosophila*. Genome Res. 11, 2101–2114.

Brenner, S.E., Chothia, C., Hubbard, T.J.P., 1998. Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. PNAS 95, 6073–6078.

Cherbas, L., Cherbas, P., 1993. The arthropod initiator - the capsite consensus plays an important role in transcription. Insect Biochem. Molec. Biol. 23, 81–90.

Cisar, C.R., Fox, J.W., Geren, C.R., 1989. Screening a brown recluse spider genomic library for the gene coding for the mammalian toxin using an oligonucleotide probe. Toxicon 27, 37–38.

Cuevas, W.A., Songer, J.G., 1993. Arcanobacterium heamolyticum phospholipase D is genetically and functionally similar to *Corynebacterium pseudotuberculosis* phospholipase D. Infect. Immun. 61 (10), 4310–4316.

Desai, A., Miller, M.J., Gomez, H.F., Warren, J.S., 1999. Loxosceles deserta spider venom induces NF-kappa B-dependent chemokine production by endothelial cells. J. Toxinol.-Clin. Toxicol. 37 (4), 447–456.

Diniz, M.R.V., Paine, M.J.I., Diniz, C.R., Theakston, R.D.G., Crampton, J.M., 1993. Sequence of the cDNA coding for the lethal neurotoxin Tx1 from the Brazilian “Armed” spider *Phoneutria nigriventer* predicts the synthesis and processing of a preprotoxin. J. Biol. Chem. 268 (21), 15340–15342.

DyrlØv Bendtsen, J., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides. J. Mol. Biol. 2004; (in press, available online).

Fernandes Pedrosa, M.F., Junqueira de Azevedo, I.L., Goncalves-de-Andrade, R.M., Van den Berg, C.W., Ramos, C.R., Ho, P.L., Tambourgi, D.V., 2002. Molecular cloning and expression of a functional dermonecrotic and haemolytic factor from *Loxosceles laeta* venom. Biochem. Biophys. Res. Commun. 298 (5), 638–645.

Fischer, D., Barret, C., Bryson, K., Elofsson, A., Godzik, A., Jones, D., Karplus, K.J., Kelley, L.A., Maccallum, R.M., Pawowski, K., Rost, B., Rychlewski, L., Sternberg, M.J., 1999. Critical assessment of fully automated structure prediction methods. Protein: Struct, Funct. Gen. 3, 209–217.

Gertsch, W.J., Ennik, F., 1983. The spider genus Loxosceles in North America Central America, and the West Indies (*Araneae, Loxoscelidae*). Bull. Am. Mus. Nat. Hist. 175, 264–360.

Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of horizontal gene transfer. Mol. Biol. Evol. 19 (12), 2226–2238.

Gomez, H.F., Miller, M.J., Waggener, M.W., Lankford, H.A., Warren, J.S., 2001. Antigenic cross-reactivity of venoms from medically important North American *Loxosceles* spider species. Toxicon 39, 817–824.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Kalapothakis, E., Peneforte, C.L., Beirao, P.S.L., Romano-Silva, M.A., Cruz, J.S., Prado, M.A.M., Guimaraes, P.E.M., Gomez, M.V., Prado, V.F., 1898. Cloning of cDNAs encoding neurotoxic peptides from the spider Phoneutria nigriventer. Toxicon 36 (12), 1843–1850.

Kalapothakis, E., Araujo, S.C., de Castro, C.S., Mendes, T.M., Gomez, M.V., Mangili, O.C., Gubert, I.C., Chavez-Olortegui, C., 2002. Molecular cloning, expression and immunological properties of LiD1, a protein from the dermonecrotic family of Loxosceles intermedia spider venom. Toxicon 40 (12), 1691–1699.

Kelley, L.A., Maccallum, R., Sternberg, M.J.E., 1999. Recognition of remote protein homologies using three-dimensional information to generate a position specific scoring matrix in the program 3D-PSSM. RECOMB 99, in: Istrail, S., Pevzner, P., Waterman, M. (Eds.), Proceedings of the Third Annual Conference on Computational Molecular Biology. The Association for Computing Machinery, New York, pp. 218–225.

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., Fitch, D.H.A., 2004. Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss. PNAS 101 (24), 9003–9008.

Kurpiewski, G., Forrester, L.J., Barrett, J.T., Campbell, B.J., 1981. Platelet aggregation and sphingomyelinase D activity of a purified toxin from the venom of *Loxosceles reclusa*. Biochim. Biophys. Acta 678, 467–476.

Maddison, D.R., Maddison, W.P., 2001. MacClade 4.02. Sinauer Associates, Inc., Sunderland, Massachusetts.

McNamara, P.J., Cuevas, W.A., Songer, J.G., 1995. Toxic phospholipases-D of *Corynebacterium pseudotuberculosis, C. ulcerans* and *Arcanobacterium haemolyticum*—cloning and sequence homology. Gene 156, 113–118.

Menez, A., 2002. Perspectives in Molecular Toxinology. Wiley, New York.

Monteiro, C.L.B., Rubel, R., Cogo, L.L., Mangili, O.C., Gremski, W., Viega, S.S., 2002. Isolation and identification of *Clostridium perfringens* in the venom and fangs of *Loxosceles intermedia* (brown spider): enhancement of the dermonecrotic lesion in loxescelism. Toxicon 40, 409–418.

Nakamura, Y., Nishio, Y., Ikeo, K., Gojobori, T., 2003. The genome stability in corynebacterium species due to lack of the recombinatorial repair system. Gene 317, 149–155.

Nakamura, Y., Iton, T., Matsuda, H., Gojobori, T., 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat. Genet. Jun 20 (e-pub).

Ramos-Cerillo, B., Olvera, A., Odell, G.V., Zamudio, F., Paniagua-Solís, J., Alagón, A., Stock, R.P. 2004. Genetic and enzymatic characterization of sphingomyelinase D isoforms from the North American fiddleback spiders *Loxosceles boneti* and *Loxosceles reclusa*. Toxicon 44:507–514.

Rees, R.S., Nanney, L.B., Yates, R.A., King Jr., L.E., 1984. Interaction of brown recluse spider venom on cell membranes: the inciting mechanisms? J. Invest. Dermatol. 83 (4), 270–275.

Rees, R.S., Gates, C., Timmons, S., Des Pres, R.M., King, J.L.E., 1988. Plasma components are required for platelet activation by the toxin of *Loxosceles reclusa*. Toxicon 26, 1035–1045.

Robertson, H.M., 1998. Two large families of chemoreceptor genes in the nematodes *Caenorhabdites elegans* and *Caenorhabdites briggsae* reveal extensive gene duplication, diversification, movement and intron loss. Genome Res. 8, 449–463.

Ramos-Cerillo, B., Olvera, A., Odell, G., Zamudio F., Paniagua-Solis, J., Alagon, A., Stock, R.P. 2004. Genetic and enzymatic characterization of sphingomyelinase D isoforms from the North Ameican fiddleback spiders *Loxosceles boneti* and *Loxosceles reclusa*. Toxicon.

Santos, A.D., Imperial, J.S., Chaudhary, T., Beavis, R.C., Chait, B.T., Hunsperger, J.P., Olivera, B.M., Adams, M.E., Hillyard, D.R., 1992. Heterodimeric structure of the spider toxin w-agatoxin IA revealed by precursor analysis and mass spectrometry. J. Biol. Chem. 267 (29), 20701–20705.

Songer, J.G., 1997. Bacterial phospholipases and their role in virulence. Trends Microbiol. 5 (4), 156–161.

Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis using Parsimony (and other methods). Sinauer Associates, Sunderland, MA.

Tambourgi, D.V., Magnoli, F.C., von Eickstedt, V.R.D., Benedetti, Z.C., Petricevich, V.L., da Silva, W.D., 1995. Incorporation of a 35-Kilodalton purified protein from *Loxosceles intermedia* spider venom transforms human erythrocytes into activators of autologous complement alternative pathway. J. Immunol. 155, 4459–4466.

Tambourgi, D.V., Magnoli, F.C., Van den Berg, C.W., Morgan, B.P., de Araujo, P.S., Alves, E.W., Dias da Silva, W., 1998. 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.

Tambourgi, D.V., Silva, M.S., Billington, S.J., Gonçalves-de-Andrade, R.M., Magnoli, F.C., Songer, J.G., Van den Berg, C.W., 2002. Mechanism of induction of complement susceptibility of erythrocytes by spider and bacterial sphingomyelinases. Immunology 107, 93–101.

Tambourgi, D.V., Fernandes Pedrosa, M.F., Van den Berg, C.W., Gonçalves-de-Andrade, R.M., Ferracini, M., Paixão-Cavalcante, D., Morgan, B.P., Rushmere, N.K., 2004. Molecular cloning, expression, function and immunoreactivities of members of a gene family of sphingomyelinases from Loxosceles venom glands. Mol. Immunol. 41, 831–840.

Tarrio, R., Rodriguez-Trelles, F., Ayala, F.J., 1998. New *Drosophila* introns originate by duplication. PNAS 95, 1658–1662.

Truett III., A.P., King, J.L.E., 1993. Sphingomyelinase D: a pathogenic agent produced by bacteria and arthropods. Adv. Lipid Res. 26, 275–291.

Van Meeteren, L.A., Fredricks, F., Giepmans, B.N.G., Fernandes Pedrosa, M.F., Billington, S.J., Jost, B.H., Tambourgi, D.V., Moolenaar, W.H., 2004. Spider and bacterial sphingomyelinse D target cellular lysophosphatidic acid receptors by hydrolyzing lysophosphatidylcholine. J. Biol. Chem. 279 (12), 10833–10836.

Viega, S.S., Gremski, W., dos Santos, V.L.P., Feitosa, L., Mangili, O.C., Nader, H.B., Dietrich, C.P., Brentani, R.R., 1999. Oligosaccharide residues of Loxosceles intermedia (brown spider) venom proteins: dependence on glycosylation for dermonecrotic activity. Toxicon 37, 587–607.

Yozwiak, M.L., Songer, J.G., 1993. Effect of *Corynebacterium pseudotuberculosis* phospholipase D on viability and chemotactic responses of ovine neutrophils. Am. J. Vet. Res. 54, 392–397.