Phylogenetic Analysis of Brown Recluse Spider Venom

Vanesa Robledo

March 17, 2023

Contents

1 Introduction **1**

2 Methodology **1**
2.1 Initial Analysis using Seaview 1
2.2 Phylogenetic Analysis in R 2

3 Results & Discussion **5**
3.1 Phylogenetic Tree Constructed Using Seaview 5
3.2 Phylogenetic Tree Constructed Using Phanghorn 8

4 References **9**

5 Session Information **9**

1 Introduction

The brown recluse spider, *Loxosceles reclusa* is one of the few venomous spiders that can be found in North America. Envenomation in mammals can result in dermonecrotic arachnidism and loxoscelism, which is the syndrome that results from envenomation. Loxoscelism involves the formation of dermonecrotic lesions and systemic conditions such as fever, fatigue, headaches, vomiting, and more. In some cases, this can result in renal failure and hemolysis. [1]

The venom of Loxosceles species contains many toxic proteins. Sphingomyelinase D (SMase D) has been found to be the main protein that causes the pathological effects [2], namely by cleavage of sphingomyelins [3].

SMase D is found in Loxosceles and Sicarius spider venom. SMase D has also been found in bacteria, and due to its similarity in spiders, has been hypothesized to be related via horizontal gene transfer. However, Binford et. al. 2005 demonstrated from sequence analysis that bacterial SMase D is not related, and it likely originated from evolution within eukaryotes. Multiple paralogs of SMaseD found within *Loxosceles arizonica* and other species venom suggest multiple versions of this gene are expressed in Loxosceles. [4]

This assignment attempts to analyze the evolutionary history of SMase D found in Loxosceles species and its homologs by constructing phylogenetic trees from homologs of SMase D. One phylogenetic tree was constructed with a set of sequences from different species, and the second was constructed focusing on homologs within spiders (*Aranae*). This assignment hypothesizes that SMase D emerged in *Loxosceles* from a gene duplication event.

2 Methodology

2.1 Initial Analysis using Seaview

2.1.1 Construction of Phylogenetic Tree

An initial set of sequences containing SMase D found in *Loxosceles reclusa* (accession number AAW56831) and homologs were given. These sequences were found in spiders (*Araneae*), other members of *Arachnida* like mites and ticks (*Acari*) and scorpions (*Scorpiones*), and the Atlantic horseshoe crab (*Limulus polyphemus*).

Table 1: Description of the sequences in the multiple sequence alignment used to create the phylogenetic tree in R. An asterisk (*) indicates a newly-added sequence to the alignment.

Sequence Name	Accession Number	Gene Name	Species Name	Size (Amino Acids)
Lsimilis_Loxtox_s1D	ANY30963	lotox	Loxosceles similis	306
Lsimilis_Loxtox_s1A	ANY30960	lotox	Loxosceles similis	306
Lsimilis_Loxtox_s5A	ANY30972	lotox	Loxosceles similis	308
Lintermedia_SMD_P2	AAP97092	sphingomyelinase P2 precursor	Loxosceles intermedia	322
Lintermedia_SMD_P1	AAP97091	sphingomyelinase P2 precursor	Loxosceles intermedia	322
Lreclusa_SMD_P1	AAT66075	sphingomyelinase D protein 1, partial	Loxosceles reclusa	279
Lreclusa_SMD_P2	AAT66076	sphingomyelinase D protein 2, partial	Loxosceles reclusa	279
Speruensis_SMDlike	ACN48988	sphingomyelinase D-like protein, partial	Sicarius peruensis	275
Ptepidariorum_SMD	XP_015929781	dermonecrotic toxin StSicTox-betaIB1i	Parasteatoda tepidariorum	320
Lspadicea_SMD_P1*	C0JAT9	sphingomyelin phosphodiesterase D	Loxosceles spadicea	273
Lspadicea_SMD_P2*	ACN48860	sphingomyelinase D-like protein, partial	Loxosceles spadicea	271
Lrufescens_SMD_P1*	AKB91150	sphingomyelinase D-like protein, partial	Loxosceles rufescens	271
Lrufescens_SMD_P2*	AKB91138	sphingomyelinase D-like protein, partial	Loxosceles rufescens	271
Lhirsuta_SMD_P1*	ACN48838	sphingomyelinase D-like protein, partial	Loxosceles hirsuta	271
Lhirsuta_SMD_P2*	C0JAT6	sphingomyelin phosphodiesterase D	Loxosceles hirsuta	273

This set of sequences aligned using the MUSCLE algorithm in Seaview 4 (Appendix 1) [5]. The Gblocks tool in Seaview was used to identify regions that align well and remove poorly-aligned and divergent regions. This multiple sequence alignment was then used to construct a phylogenetic tree using the Neighbor-Joining method and with 100 bootstrap replicates in Seaview (Figure 1/Appendix 2).

2.2 Phylogenetic Analysis in R

2.2.1 Creating Multiple Sequence Alignment

After the first phylogenetic tree of the SMase D homologs was constructed in Seaview, several distant sequences were removed and more homologs found from within the *Loxosceles* genus were added. Sequences that were not in the *Araneae* suborder were removed to more closely examine evolutionary events related to SMase D in recluse spiders.The homolog from *P. tepidaiorum* was kept as an outgroup to the sphingomyelinase proteins.

The sequences removed from the initial set are as follows:

- ABD73957_SMDlike_Iscapularis
- XP_022646068_SMDlike_Vdestructor
- XP_013777636_SicToxlike_Lpolyphemus
- API81378_venomtox_Hlepturus

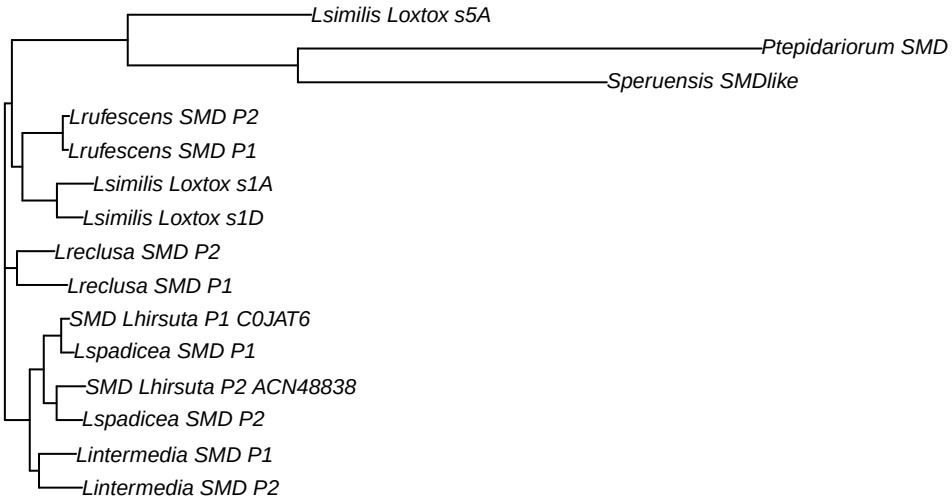
Table 1 summarizes all of sequences included in the multiple sequence alignment. This new set of sequences was then aligned using the same methods as the first set of sequences (Appendix 3).

2.2.2 Constructing Neighbor-Joining Tree in R

The new multiple sequence alignment was loaded and analyzed using the `phangorn` package in R [6].

```
# BIOL 365 Assign3 Phylogeny in R
# Vanesa Robledo

# Load required packages
library(Biostrings)
library(phangorn)

# Load MSA
Lotox_phydat <- read.phyDat("Loxtox_sequences_edited.fst", format = "fasta", type ="AA")

# Construct a neighbor-joining tree:
```

```
# Create distance matrix
Lotox_dm <- dist.ml (Lotox_phydat, model = "JTT")

# Construct NJ tree & plot
Lox_tree <- NJ(Lotox_dm)
plot.phylo(Lox_tree, cex = 0.7)
```



2.2.3 Testing Evolutionary Models

The `modelTest()` function was used to test three evolutionary models: JTT, LG, and WAG with the constructed neighbor-joing tree. The quality of the models were assessed using Akaike Information Criterion (AIC) and the best model was selected based on having the lowest AIC.

```
# Use modelTest() to test 3 evolutionary models:
mt <- modelTest(Lotox_phydat, model = c("JTT", "LG", "WAG"))
```

```
## Model        df  logLik   AIC      BIC
##          JTT 29 -3562.306 7182.612 7285.76
##        JTT+I 30 -3508.467 7076.934 7183.639
##     JTT+G(4) 30 -3479.75 7019.501 7126.206
##   JTT+G(4)+I 31 -3479.082 7020.163 7130.425
##           LG 29 -3532.836 7123.672 7226.82
##         LG+I 30 -3481.507 7023.014 7129.719
##      LG+G(4) 30 -3449.514 6959.028 7065.733
##    LG+G(4)+I 31 -3449.126 6960.251 7070.513
##          WAG 29 -3511.84 7081.68 7184.828
##        WAG+I 30 -3470.28 7000.56 7107.265
##     WAG+G(4) 30 -3449.854 6959.708 7066.413
##   WAG+G(4)+I 31 -3449.062 6960.124 7070.386
```

```
mt
```

```
##          Model df    logLik      AIC         AICw     AICc        AICcw      BIC
## 1          JTT 29 -3562.306 7182.612 9.935980e-50 7190.210 1.460351e-49 7285.760
## 2       JTT+I 30 -3508.467 7076.934 8.807826e-27 7085.092 9.785670e-27 7183.639
## 3     JTT+G(4) 30 -3479.750 7019.501 2.608483e-14 7027.659 2.898076e-14 7126.206
## 4  JTT+G(4)+I 31 -3479.082 7020.163 1.872883e-14 7028.903 1.555288e-14 7130.425
## 5           LG 29 -3532.836 7123.672 6.251138e-37 7131.270 9.187673e-37 7226.820
## 6        LG+I 30 -3481.507 7023.014 4.502786e-15 7031.172 5.002686e-15 7129.719
## 7      LG+G(4) 30 -3449.514 6959.028 3.530379e-01 6967.186 3.922321e-01 7065.733
## 8   LG+G(4)+I 31 -3449.126 6960.251 1.915571e-01 6968.991 1.590737e-01 7070.513
## 9          WAG 29 -3511.840 7081.680 8.208622e-28 7089.279 1.206470e-27 7184.828
## 10       WAG+I 30 -3470.280 7000.560 3.382968e-10 7008.718 3.758545e-10 7107.265
## 11    WAG+G(4) 30 -3449.854 6959.708 2.513036e-01 6967.866 2.792033e-01 7066.413
## 12 WAG+G(4)+I 31 -3449.062 6960.124 2.041014e-01 6968.864 1.694908e-01 7070.386
```

```
# Find the index of the models with the lowest AIC value using which.min()
which.min(mt$AIC)
```

```
## [1] 7
```

```
# Set the best model (lowest AIC) to object 'bestmodel'
env = attr(mt, "env")
bestmodel <- mt$Model[which.min(mt$AIC)]
```

The best model is **LG+G(4)+I**. Since the model includes gamma distrubtion (G(4)) and invariant sites (I), both the gamma and invariant sites parameters must be optimized (`optGamma` and `optInv` set to `TRUE`) in `optim.pml()`.

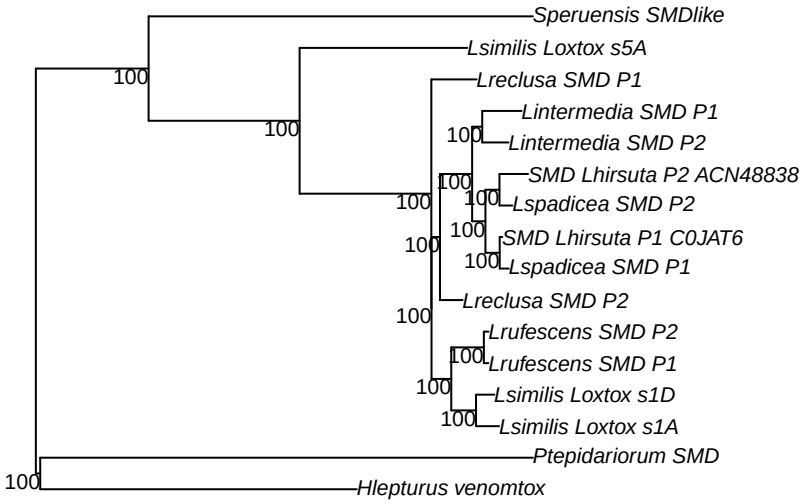
```
# Fit model to maximum likelihood using stochastic algorithm
fitStart = eval(get(bestmodel, env), env)
fit = optim.pml(fitStart, rearrangement = "stochastic", optGamma=TRUE, optInv=TRUE, bestmodel)
```

```
fit
```

```
## model: LG+G(4)+I
## loglikelihood: -3446.673
## unconstrained loglikelihood: -1351.86
## Proportion of invariant sites: 0.06157526
## Discrete gamma model
## Number of rate categories: 4
## Shape parameter: 1.1253
## Rate matrix: LG
```

```
# Add bootstrap values:
bs = bootstrap.pml(fit, bs = 100, optNnni = TRUE)
```

```
# Plot the copy of tree with bootstrap values
plotBS(midpoint(fit$tree), bs, p = 50, type="p", cex = 0.7)
```



3 Results & Discussion

3.1 Phylogenetic Tree Constructed Using Seaview

This phylogenetic tree included homologs of SMase D outside of *Loxosceles* species. In general, this tree shows that paralogs in *Loxosceles* species are more closely-related to each other than to other species.

Some of the branches are well supported (i.e. have a bootstrap confidence level of over 70):

- Paralogs of Loxtox in *L. similis* and SMase D in *L. intermedia*,
- *Sicarius* branching out from *Loxosceles*
- *H. lepturus* branching out from the *Sicariius* family, and
- *I. scapularis*, *V. destructor*, and *L. polyphemus* branching out from the rest of the *Aranae*.

The phylogeny suggests recent gene duplications within *Loxosceles*, although it lacks support. Loxtox paralogs in *L similis* branch out from the rest of the family but there is no support for it. There is also a lack of support for the branches separating the *L. intermedia* paralogs and *L. reclusa* paralogs, and the paralogs of *L. reclusa* also lack support.

Not all of representative sequences of the branches of the phylogeny agree with expected species phylogeny (Figure 3) [7] . Most of them agree, but there are notable exceptions. The scorpion (*H. lepterus*) is considered more closely-related to the rest of the spiders than the common house spider (*P. tepidariorum*), even though scorpions (*Scorpiones*) diverged from *Aranae* earlier than *P. tepidariorum*. The Atlantic horshoe crab (*L. polyphemus*) is also grouped with the mite (*V. destructor*) and tick (*Ixodida*), despite taxonomic evidence that the horseshoe crab diverged from the rest of *Arthropoda* earlier from mites and ticks, which are in the *Arachnida* family along with the scorpion and spiders.

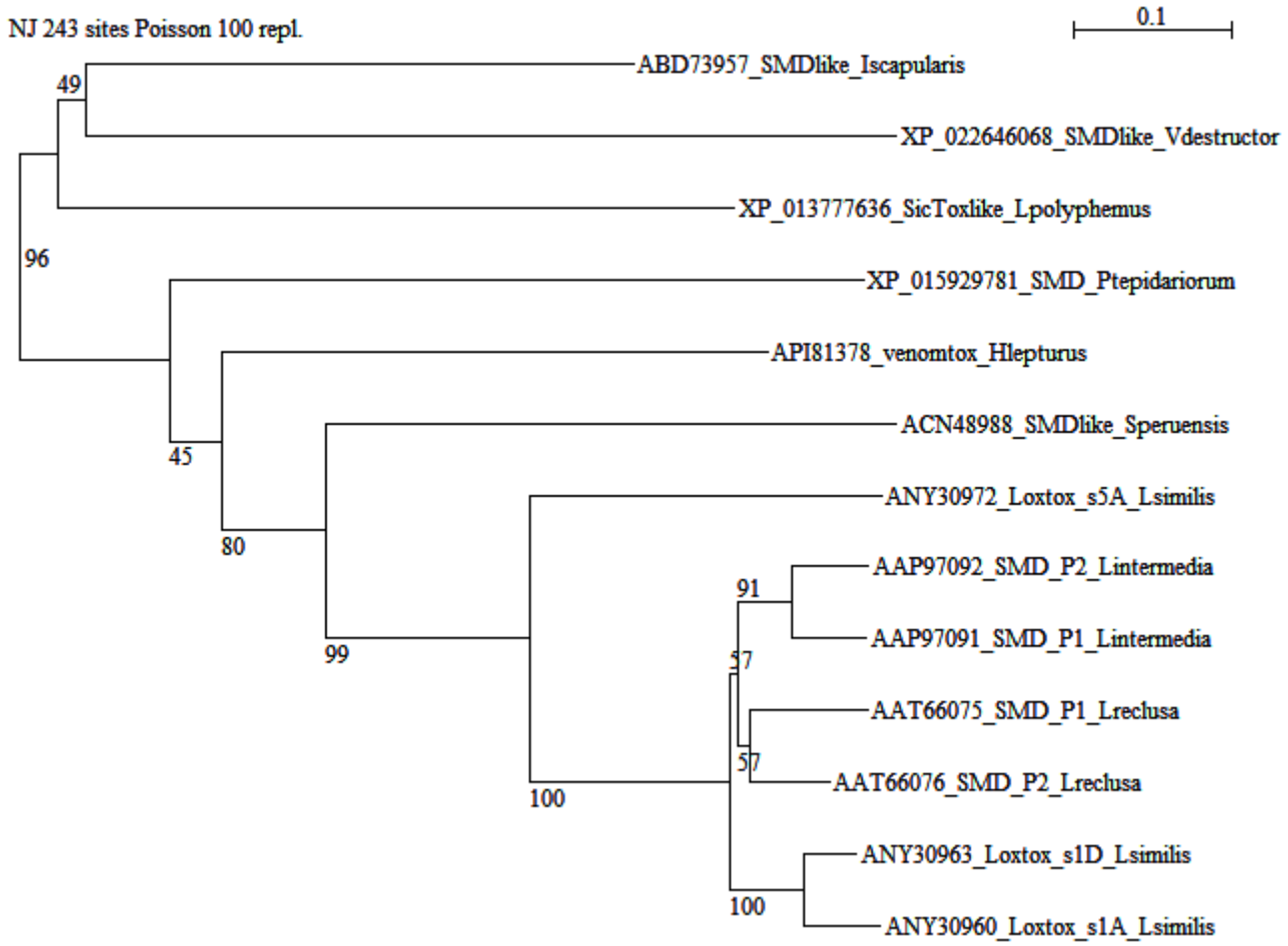


Figure 1: Phylogenentic tree of SMase D homologs present in Loxosceles species and homlogs present in other species constructed using Seaview. The Neighbor-Joining method with Poisson distribution was selected using 100 bootstrap replicates. The numbers on the branches indicate the strength of the support of each relationship.

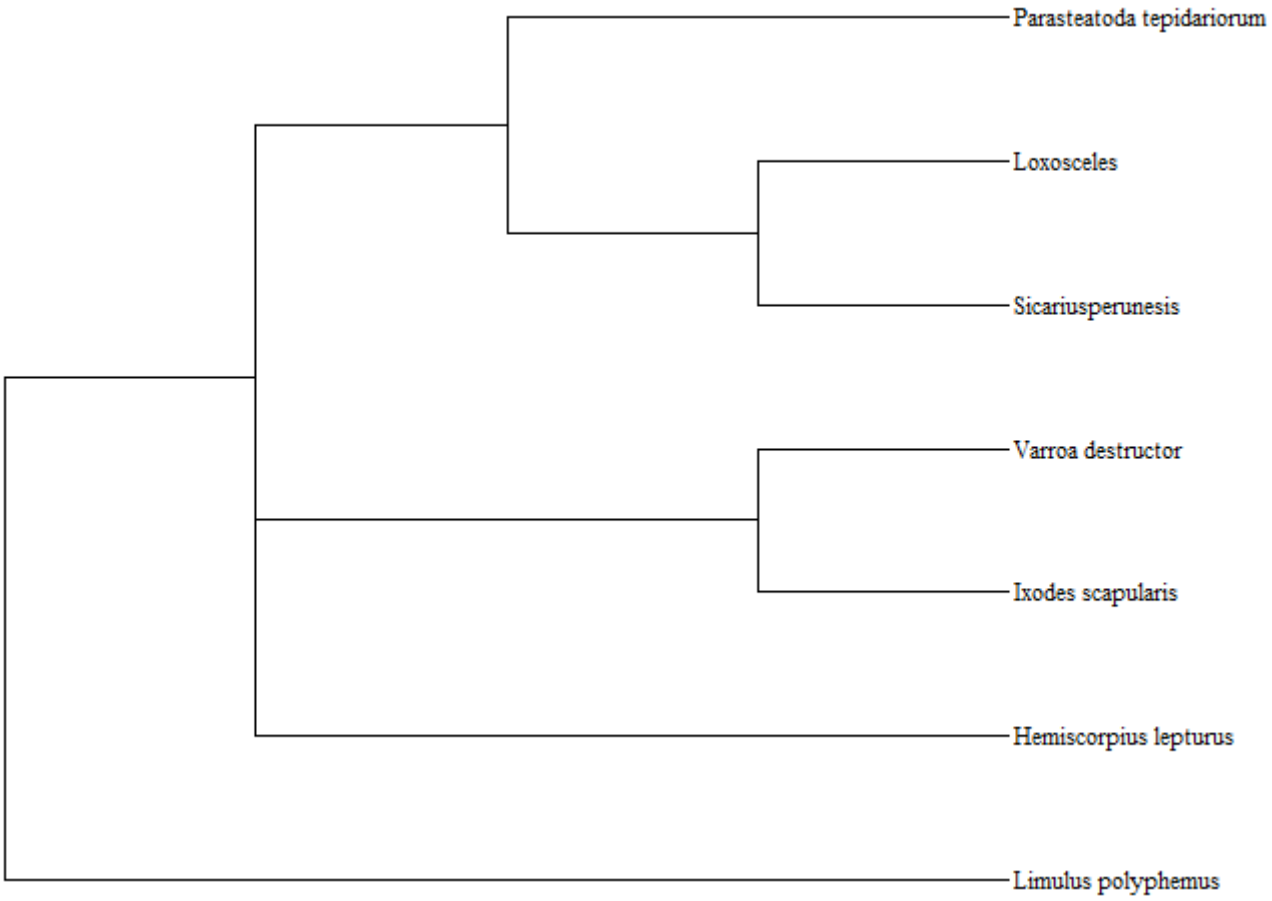


Figure 2: Phylogeny describing relationship between the species included in the sequences as described by NCBI Taxbrowser. This was created using Newick format and viewed through Seaview.

3.2 Phylogenetic Tree Constructed Using Phanghorn

All of the branches in phylogenetic tree created using `phanghorn` are well supported (i.e. bootstrap confidence values are 100). This phylogenetic trees with expected taxonomy within the Sicariidae phylogeny [8].

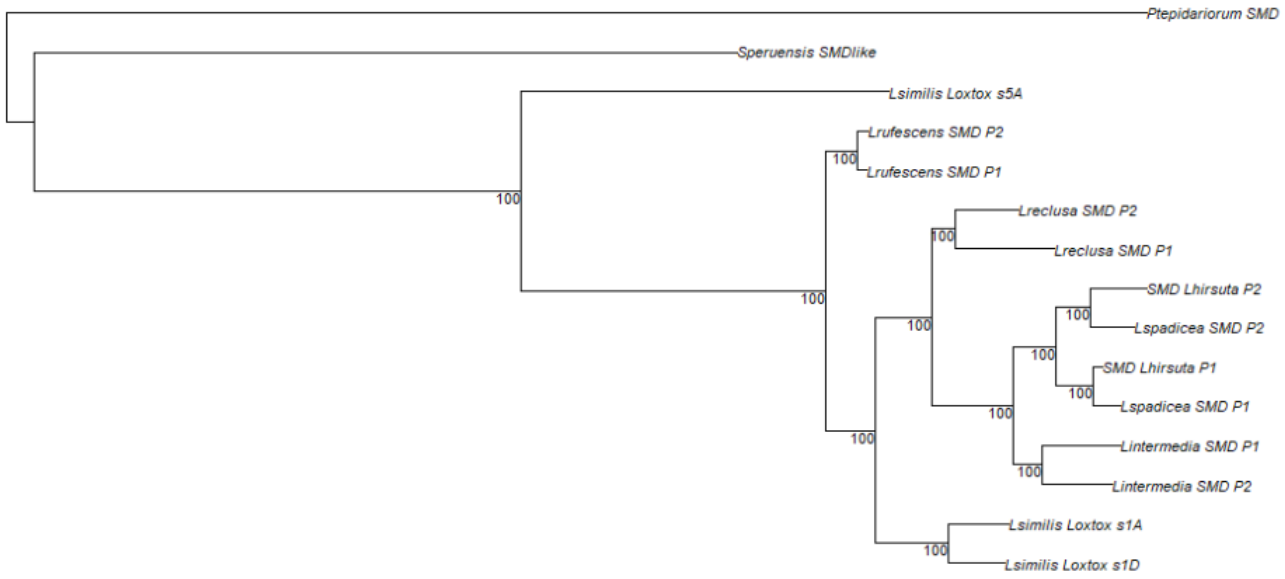


Figure 3: Phylogenetic tree constructed using the `phangorn` package in R of spider toxins. The numbers on the branches represent bootstrap confidence level.

The phylogenetic tree also agree with expected Loxtox gene phylogeny [9] - *L similis* loxtox s5A is more distantly related than loxtox s1A and s1D. Interestingly, *L similis* loxtox s5A is more distantly related to loxtox s1A and s1D, which are more closely related to *L reclusa*, *L hirsuta*, *L spadicea*, and *L intermedia* than *L refuscens* is. This suggests either convergent evolution or a gene duplication event when *L similis* diverged from the other Loxosceles species, although more loxtox sequences should be included to support either hypothesis.

The sequences found within Loxosceles are grouped together and more related to each other than the Sicarius SMaseD-like protein, suggesting SMase D emerged after Loxosceles and Sicarius diverged.

There is evidence of possible genetic duplication events within Loxosceles species. The sequences of genes in *L. reclusa*. *L. refuscens*, and *L. intermedia* are more closely related within the species than to other species, suggesting paralogy and gene duplication events within the species.

L. hirsuta and *L. spadicea* are considered each other's closest relatives and both closely-related to *L. intermedia* [9], which is also reflected in this phylogeny. The sequences of *L. hirsuta* and *L. spadicea* are more closely related to the other than within species. Shingomyelin phosphodiesterase D found in *L. hirsuta* is suggested to be homologous with sphingeomyelinase D-like protein, partial found in *L. spadicea* and the sphingomyelinase D-like protein, partial protein found in *L. hirsuta* is suggested to be homologous to Shingomylein phosphodiesterase D. This may suggest a gene duplication event in that occurred before *L. hirsuta* and *L. spadicea* diverged that led to the evolution of sphingomyelin phosphodiesterase D.

Binford et. al. 2009 constructed a phylogeny of SMase D in Sicariid spider venom and proposed a gene group, *SicTox* [10]. They propose two major clades within Sicariidae: the α clade, which contains species that produces SMase D and have dermonecrotic activity, and the β clade that has no or reduced SMase D and contains paralogs from Sicarius. The α clade includes *L. refuscens*, *L. reclusa*, *L. intermedia*, *L. hirsuta*, and *L. spadicea*. The β cladde includes *S. peruensis* and paralogs of SMD from *L. hirsuita* and *L. intermedia*. This phylogenetic tree agrees with these groupings: *S. peruensis* is more divergent from the rest of Loxosceles and the sphingomyelinase D protein family is together in its own clade.

Overall, this phylogenetic tree supports the hypothesis of gene duplication within Loxosceles. Multiple gene duplication events are implied and supported by other phylogenetic studies of SMase D.

4 References

1. Chaves-Moreira, D., Senff-Ribeiro, A., Wille, A.C.M., Gremski, L.H., Chaim, O.M. and Veiga, S.S., 2017. Highlights in the knowledge of brown spider toxins. *Journal of Venomous Animals and Toxins including Tropical Diseases*, 23.
2. Tambourgi, D.V., Magnoli, F.C., van den Berg, C.W., Morgan, B.P., de Araujo, P.S., Alves, E.W. and Da Silva, W.D., 1998. Sphingomyelinases in the venom of the SpiderLoxosceles intermedia are responsible for both dermonecrosis and complement-dependent Hemolysis. *Biochemical and Biophysical Research Communications*, 251(1), pp.366-373.
3. Lopes, P.H., Berg, C.W.V.D. and Tambourgi, D.V., 2020. Sphingomyelinases D from Loxosceles spider venoms and cell membranes: action on lipid rafts and activation of endogenous metalloproteinases. *Frontiers in Pharmacology*, 11, p.636.
4. Binford GJ, Cordes MH, Wells MA. Sphingomyelinase D from venoms of Loxosceles spiders: evolutionary insights from cDNA sequences and gene structure. *Toxicon.* 2005 Apr;45(5):547-60. doi: 10.1016/j.toxicon.2004.11.011. PMID: 15777950.
5. Guoy, M., Guindon, S. and Gascuel, O., 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building.
6. Schliep K (2011). "phangorn: phylogenetic analysis in R." Bioinformatics, 27(4), 592–593.
7. Schoch CL, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020: baaa062. PubMed: 32761142 PMC: PMC7408187.
8. Binford, G.J., Callahan, M.S., Bodner, M.R., Rynerson, M.R., Núñez, P.B., Ellison, C.E. and Duncan, R.P., 2008. Phylogenetic relationships of Loxosceles and Sicarius spiders are consistent with Western Gondwanan vicariance. *Molecular Phylogenetics and Evolution*, 49(2), pp.538-553.
9. Dantas, A.E., Carmo, A.O., Horta, C.C.R., Leal, H.G., Oliveira-Mendes, B.B.R., Martins, A.P.V., Chávez-Olórtegui, C. and Kalapothakis, E., 2016. Description of Loxtox protein family and identification of a new group of Phospholipases D from Loxosceles similis venom gland. *Toxicon*, 120, pp.97-106.
10. Binford, G.J., Bodner, M.R., Cordes, M.H., Baldwin, K.L., Rynerson, M.R., Burns, S.N. and Zobel-Thropp, P.A., 2009. Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase D in sicariid spider venoms. *Molecular Biology and Evolution*, 26(3), pp.547-566.

5 Session Information

All of the output in this assignment was produced under the following conditions:

- R version 4.2.2 (2022-10-31 ucrt), `x86_64-w64-mingw32/x64 (64-bit)`
- Running under: Windows 10 x64 (build 19045)
- Matrix products: default
- Base packages: stats4, stats, graphics, grDevices, utils, datasets, methods, base
- Other attached packages: phangorn_2.11.1, ape_5.6-2, Biostrings_2.66.0, GenomeInfoDb_1.34.6, XVector_0.38.0, IRanges_2.32.0, S4Vectors_0.36.1 , BiocGenerics_0.44.0, kableExtra_1.3.4, knitr_1.42

- Loaded via a namespace (and not attached): xfun_0.36, bslib_0.4.2, lattice_0.20-45, colorspace_2.1-0, generics_0.1.3vctrs_0.5.2, htmltools_0.5.4, viridisLite_0.4.1, yaml_2.3.7, rlang_1.0.6, jquerylib_0.1.4, glue_1.6.2, GenomeInfoDbData_1.2.9 lifecycle_1.0.3, stringr_1.5.0, zlibbioc_1.44.0, munsell_0.5.0, rvest_1.0.3, codetools_0.2-18, evaluate_0.20, fastmap_1.1.0, parallel_4.2.2, highr_0.10, Rcpp_1.0.10, scales_1.2.1, cachem_1.0.6, webshot_0.5.4, jsonlite_1.8.4, systemfonts_1.0.4, fastmatch_1.1-3, digest_0.6.31, stringi_1.7.12, grid_4.2.2, bitops_1.0-7, quadprog_1.5-8, cli_3.6.0, tools_4.2.2, magrittr_2.0.3, sass_0.4.5, RCurl_1.98-1.9, crayon_1.5.2, pkgconfig_2.0.3, Matrix_1.5-3, xml2_1.3.3, rmarkdown_2.20, svglite_2.1.1, httr_1.4.4, rstudioapi_0.14, R6_2.5.1, igraph_1.3.5, nlme_3.1-160, compiler_4.2.2