Silks Specific Transcripts Across the Spider Phylogeny

Christian C. Ramírez [1], Sandra Correa-Garhwal [2], Dimitri Forero [3]

[1] Laboratorio de Entomología, Departamento de Biología, Pontificia Universidad Javeriana, Bogotá, Colombia

[2] Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, California, United States of America

[3] Laboratorio de Entomología, Departamento de Biología, Pontificia Universidad Javeriana, Bogotá, Colombia

Abstract

Spiders (Order Araneae) produce silk used for everyday tasks ranging from prey-capture and immobilization to reproduction and safety draglines. Spider silks are mainly composed of proteins called spidroins, which are members of the spidroin gene family. Spider silk is not only one of the toughest materials on earth, it is also a known non-immunogenic, flexible, and lightweight. Recently, silk specific transcripts (SSTs) were described for cob-web weavers (Theridiidae) increasing the interest in studying their role in silk composition and physical properties to allow the production of synthetic spider silk. We evaluated the presence of SSTs on three spider species: *Acanthoscurria geniculata* (Theraphosidae), *Nephila clavipes* (Araneidae) and *Stegodyphus mimosarum* (Eresidae) to investigate if these SSTs are conserved across families not closely related or if they are lineage specific. We found 148 SSTs to be present in *A. geniculata*, 220 in *N. clavipes* and 137 in *S. mimosarum*. Moreover, we found 121 SSTs to be present in all species. Most SSTs shared by all species are involved in molecular functions such as oxidoreductase activity, hydrolase activity and oxidation-reduction processes. Our results highlight the importance of not only spidroins but also of SSTs in spider silk production. We propose SSTs play an important role in silk production and their study can shed light into mechanisms to the production of synthetic silk with matching mechanical properties.

Introduction

Spider silk is the strongest natural fiber known to date (Blamires et al. 2017). Spider silk is essential for a spider's survival, and spiders are known to use silk throughout their lives (Lewis 2006). Spider silk is an extracellular proteinous fiber produced in specialized abdominal glands that are in turn connected to spinnerets (Coddington 1989). Silk is used by spiders in the construction of prey-catching webs, egg case production, prey wrapping, among others (Xu & Lewis 1990, Craig & Riekel 2002, Hinman & Lewis 1992, Hayashi et al. 1999, Lewis 2006, Hayashi & Lewis 2001). The biomechanical properties of spider silk have been of great interest to the scientific community (Blackledge et al. 2011). Given that harvesting spiders for mass production of silk is impractical, the production of synthetic fibers that retain spider silk's outstanding mechanical properties of has been the focus of biologists and engineers for the past 20 years (Blamires et al. 2017; Tokoreva et al. 2013).

Spider silk is mainly composed of proteins called spidroins (spider silk fibroins, Hinman & Lewis 1992), which belong to a single gene family (Guerette et al. 1996). The core region of spidroins is highly repetitive and dominated by the amino acids alanine, serine and glycine (Xu & Lewis 1990, Blackledge 2012). The arrangement and frequency of these repeating amino acids determines the secondary and tertiary structures and is thought to contribute to mechanical properties specific to each silk type (Hayashi et al. 1999, Blackledge 2012, Lombardi & Kaplan 1990, Hayashi & Lewis 2001, Xu & Lewis 1990, Craig & Riekel 2002, Guinea et al. 2012, Xu et al. 2014).

Silk has many properties of interest, for example, major ampullate silk (dragline) has high strength and high extensibility, which makes it extremely though; flagelliform silk has an extraordinary extensibility while having good strength (Hayashi & Lewis 2001; Blamires et al. 2017). Moreover, spider silks are lightweight and non-immunogenic (Allmeling et al. 2006, Singha et al. 2012; Blamires et al. 2017). These properties have driven the aims to produce synthetic spider silk for industrial, textile, biomedical, and biomimetic applications. These applications include lightweight and high resistance bulletproof vests, weather resistant clothes artificial ligaments and tendons to repair damage such as ruptures (Allmeling et al. 2006, Singha et al. 2012).

In recent years, next generation sequencing has been used to identify the sequence composition of spider silks and analyze what makes spiders silk so unique and how this relates to the physical properties (Babb et al. 2017). Deep sequencing and the construction of *de novo* transcriptomes have been proven useful in identifying specific transcripts and genes associated with different silk specific tissues (Clarke et al. 2014; Sanggaard et al. 2014). Clarke and colleagues identified silk-specific transcripts as strong candidates for future studies on the building blocks of spider silk and how to take that information into the production of fully functional fibers (Clarke et al. 2014).

Clarke et al (2014), identified 647 silk specific transcripts (SSTs) in the silk glands of the Western black Widow, *Latrodectus hesperus* (Theridiidae). It was shown that 70% of the silk gland expression was from SSTs and included various functions, highlighting the importance of these transcripts. Despite the potential importance of these SSTs, no further search of these transcripts has been carried out in other spider species. Therefore, here we evaluated the presence of these transcripts in spiders from different families that use silk differently to investigate how SSTs numbers differs across the spider phylogeny. We searched the publicly available transcriptomes of *Acanthoscurria geniculata* (Theraphosidae) *Nephila clavipes* (Araneae), and *Stegodyphus mimosarum* (Eresidae) for the presence of SSTs. We expected that spiders using same silks and thus having the same silk gland types would share a higher number of the SSTs. Likewise, we expected to find fewer SSTs on early diverging families that use fewer silk types and have fewer silk gland types.

Materials and Methods

Acanthoscurria geniculata, Nephila clavipes, and *Stegodyphus mimosarum* were selected to evaluate the presence of SSTs (Silk Specific Transcripts). These three species were selected based on the type of data available on NCBi. All three species have transcriptome shotgun assemblies (TSA) and genomes. They were also chosen because they belong to different families allowing for the study of SSTs variation across the spider phylogeny. Transcriptome shotgun assemblies (TSA) were retrieved from NCBi online database for the analysis (see Table 1). Each transcriptome was downloaded in FASTA format and uploaded to the Pontificia Universidad Javeriana informatics cluster –ZINE—(Garrels 2010).

Table 1 **Accession numbers** All the accession numbers for the information used during the study are given (Babb et al 2017; Sanggaard et al. 2014)

	A. geniculata	*N. clavipes*	*S. mimosarum*
Bio project NCBi	PRJNA222716	PRJNA356433	PRJNA222714
TSA	GAZS00000000	GFKT00000000	GAZR00000000
Sequencing	SRA106216	SRP095945	SRA106215

Silk specific transcripts (SSTs) from the cob-web weaver *L. hesperus* were obtained from Clarke et al. 2014 and a total of 647 SSTs were used (Clarke et al. 2014). SSTs were uploaded to the Pontificia Universidad Javeriana computing cluster –ZINE--. A database was created using SSTs protein sequences using BLAST command makeblastdb (Camacho et al. 2013). Subsequently BLASTX searches were used to search each transcriptome assembly for SSTs. An e-value of e^-20 was to reduce the number of possible mismatches and errors on the identification of SSTs. The e-value is a parameter used to describe the number of hits that can be expected by chance during the BLAST run (Kerfeld & Scott 2011; Altschul & Koonin 1998). For the e-values, the bigger the number, the more it can happen by chance, the closer the number is to zero the more significant it is (Kerfeld & Scott 2011). The e-value was chosen based on previous studies using BLAST searches were values ranging from e^-5 to e^-50 were used (Clarke et al. 2014; Tokareva et al. 2013).

Results from BLASTX searches gave the e-value, the number of mismatches and the length of the alignment all of which were taken into the account for choosing the best match. BLASTX results were retrieved and analyzed individually for each specie using Microsoft Office Excel. We obtained several matches for each SST likely due to the way the transcriptomes are fragmented for each spider which are not necessarily the same as for *Latrodectus hesperus*. For each spider the best match was selected based on the e-value, the number of mismatches and the best aligned sequences. Each selected transcript was further examined to confirm proper alignment using Galaxy (Giardine 2005; https://usegalaxy.org/). Galaxy allows the evaluation of each alignment by visually comparing both sequences (Fig 1).

Identified SSTs from each species were visualized with a Venn diagram generated using the online tool Gene List Venn Diagram (Fig 3; http://genevenn.sourceforge.net/). This was done to evaluate which of the SSTs were shared between the different families included in this study.

To investigate the function of all SSTs, we obtained Gene Ontology (GO) annotation from Clarke et al. (2014) which were assigned using UniProt. GO term description was retrieved using the online tool AmiGO 2 (http://amigo.geneontology.org/amigo/landing). The top 10 functions for the SSTs were reported using a bar graph made using Microsoft Office Excel (Fig 5).

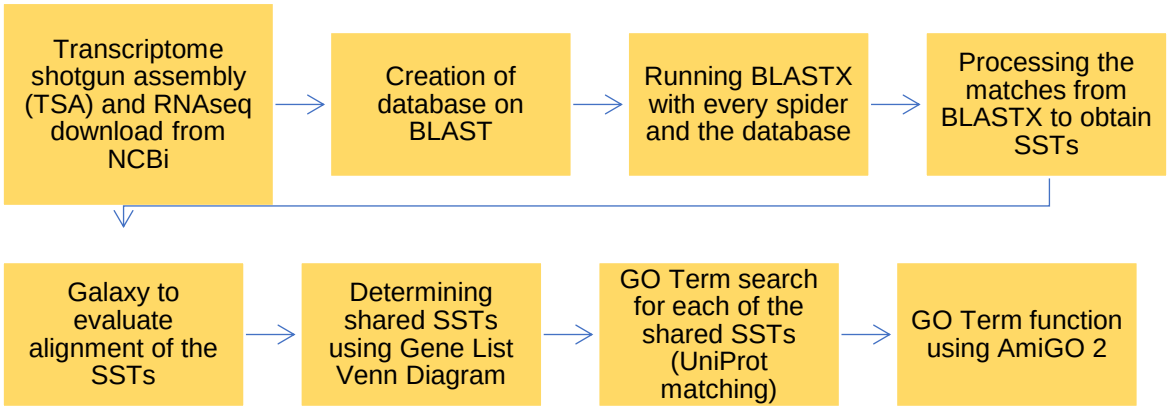

Figure 1 **Flowchart of the process realized to process the data for each of the spiders.** Process to retrieve the SSTs homolog matches from each of the spiders is detailed step by step.

Results and discussion

A total of 42,665 matches were found from BLASTX searches for the three species combined. From these matches 1,463 were from *S. mimosarum*, 3,076 from *A. geniculata*, and 38,126 from *N. clavipes* (Fig 2). Differences in the number of matches found between the three species could be due to the quality and quantity of available information used. For example, *N. clavipes* has about eight times as much genetic information available in NCBI and it is also more closely related to *L. hesperus* compared to the other species which could account for the higher number of identified SSTs (Figures 3 & 5). Although the number of matches varied between the three species after sorting and selecting the best homolog for each SSTs the number of SSTs for each specie was similar. For example, after the first selection the number of SSTs in *S. mimosarum* went down to 137, 178 in *A. geniculate*, and 220 in *N. clavipes* (Fig 2). The fact that not all SSTs reported for *L. hesperus* were identified as present in our focal species, it is likely that some of these SSTS are species-specific. Tissue-specific studies are needed to evaluate the extend of specie-specific SSTs.

Based on the close phylogenetic relationship between Araneidae and Theridiidae and the fact that both use similar silk types, we expected to find a higher number of the 647 SSTs reported for *L. hesperus* in *N. clavipes*. This was supported in our results with N. clavipes having the highest number of SSTs (220). We also found *N. clavipes* to share the most (60) SSTs with *L. hesperus* when compared to S. *mimosarum* (1) and *A. geniculata* which are shown by the SSTs not shared between the spiders studied as the original SSTs were taken from *L. hesperus* (Figure 3). This could be due to both species having more specialized silk glands (i.e. aggregate silk glands) than the other species. Additionally, it suggests that some of those silk specific transcripts could be produced in silk glands found present in *L. hesperus* and *N. clavipes* but not in *S. mimosarum* and *A. geniculata*.

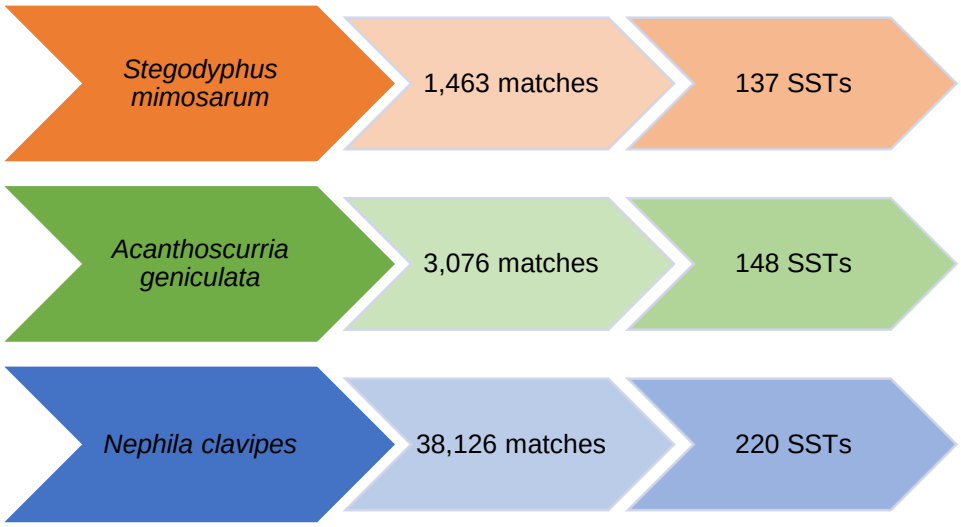

Figure 2 **Flowchart of the results from BLASTX and the SSTs obtained from those matches.** The results where filtered after being processed using BLASTX to retrieve the best homolog for each of the hits considering the e-value and how good was the alignment based on the number of mismatches, the length of the alignment between both sequences and the quality.

We found 121 SSTs to be shared among our focal species, 22 between *N. clavipes* and *A. geniculata*, 15 between *S. mimosarum* and *N. clavipes*, and none between *S. mimosarum* and *A. geniculata* (Figure 3). Our results suggest that regardless of the phylogenetic placement or how many different silk types a spider produces, there are genes that have been conserved for 203.4-327.8 Mya thus, suggesting an important role in silk production in general (Fig 5).

Recent phylogenetic studies show *N. clavipes* (Araneidae) being more closely related to *L. hesperus* (Theridiidae) than to *S. mimosarum* (Eresidae) and *A. geniculata* (Theraphosidae) (Fig 5; Fernandez et al. 2018; Wheeler et al. 2016), suggesting that a higher number of SSTs should be shared between *N. clavipes* and *S. mimosarum* than between *N. clavipes* and *A. geniculata*. Our findings show the contrary, with *N. clavipes* sharing more SSTs with A. *geniculata* than with *S. mimosarum* (22 vs. 15; Figure 3). This could be due to data from *A. geniculata* being more complete than the one from *S. mimosarum*. This is consistent with *A. geniculata* having more matches to SSTs than *S. mimosarum* (Figure 2). Another possible explanation is that some SSTs are ancestral but were lost in *S. mimosarum*. To test this, evaluation of the absence/presence of these SSTs in multiple spider species across the phylogeny would be needed.

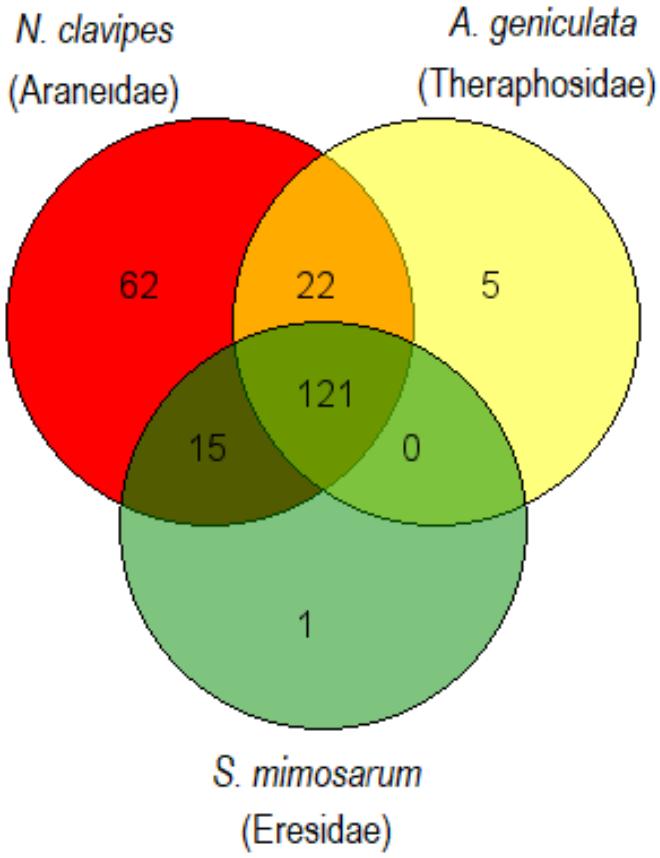

Figure 3 **SSTs relationship between the three spider species.** Venn diagram with the results from the BLASTX run. *Nephila clavipes* (red) with a total of 220 SSTs, *Acanthoscurria geniculata* (yellow) with 148 SSTs and *Stegodyphus mimosarum* (green) with 137 SSTs. 121 SSTs shared between the three species.

Given that 121 SSTs were shared among the spider families studied it is interesting to study the functions get an idea of what they are used for and how they can be shaping the properties of spider silk. Based on the GO term analysis, we found 101 of the SSTs reported in the study to have GO terms with them. Twenty of the shared SSTs did not have any GO term reported. SSTs functions are mainly classified into cellular components, biological processes or molecular functions categories although most of them were underrepresented with just one (1) match for each function. We show the top 10 most represented functions reported for SSTs shared across species (Fig 4).

There is a wide range of functions covered by SSTs. SSTs shared by the four spider families have functions mainly related with oxygen, they are enriched in oxidoreductase activity among all the top 10 functions found (37 matches) oxidation-reduction processes (21 matches), and metal ion binding (14 matches). They also have a notable number of matches with peptidase, transferase, catalytic and hydrolase activity. The function of SSTs gives an idea of their role in silk production and give a new focus of attention for future studies in how these functions might help give spider silk some of its properties.

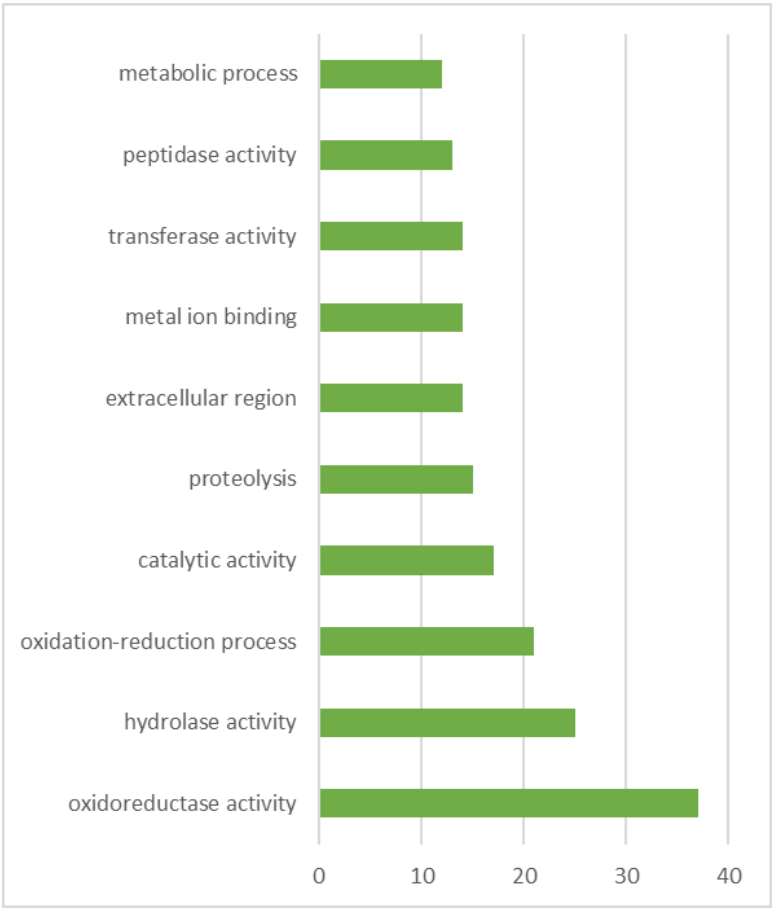

Figure 4 **GO term representation for the silk specific transcripts (SSTs) shared between spider families** Top 10 of the functions represented by the SSTs found to be shared between spider families. More than 130 functions were reported, all of them correspond to biological processes, molecular function or cellular components.

To date, *L. hesperus* is the only spider species for which SSTs have been identified (Clarke et al. 2014). In this study we add three more species and show how some SSTs are conserved. This is then the first-time silk specific transcripts shared between four families found across the spider phylogeny are reported. The first report for spiders is from almost 400 Mya (Fernandez et al. 2018). We propose that these transcripts should have appeared in the period before the split of Araneomorphae and Mygalomorphae 203.4 – 327.8 Mya (Fernandez et al. 2018; Wheeler et al. 2016) (Fig 5). This suggests that these 121 SSTs are very important for silk production in general whether the spider has one or seven silk glands types. Future studies could investigate whether these SSTs influence silk's mechanical properties. Moreover, it will shed light into the fundamental structure of spider silk which would give new insights and information on the production of synthetic spider silk.

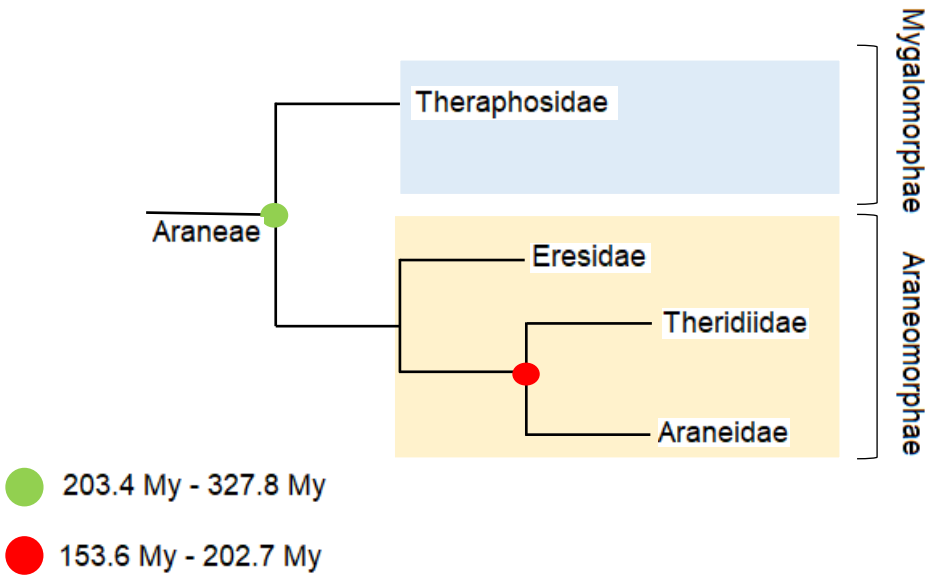

Figure 5 **Relationship between the spiders analyzed in the study.** Phylogeny based on Fernandez et al. 2018 and Wheeler et al. 2016 showing the relationships between the different families used in this study. In colored squares are shown the two sub orders of Araneae and the families included in this sub orders. The colored circles indicate approximate divergence time for different families. in time in which some groups diverged.

Conclusion

Spider silk have gained the attention of many due unmatched mechanical properties and the potential for uses for biomedical applications. We found Theridiidae silk specific transcripts to be present in different families across the spider phylogeny (Figure 2). Specifically, we found 121 SSTs that are present in our focal species suggesting these genes have conserved for ~300 Mya (Figure 3). Our findings propose potential candidates for studies analyzing the composition of spider silk not just on spidroins but on silk-specific transcripts that likely play an important role in fiber processing.

Acknowledgements

We thank Alexander Martin Herrera from Pontificia Universidad Javeriana for the help with the computer analyses done in the cluster. We thank to Dr. Paul Babb from the Perelman School of Medicine at the University of Pennsylvania and Dr. Thomas Clarke from the Department of Biology at Washington & Lee University for providing information regarding data downloading. This work was supported by the Pontificia Universidad Javeriana.

References

1. Allmeling, C., Jokuszies, A., Reimers, K., Kall, S. & Vogt, M. (2006). Use of spider silk fibres as an innovative material in a biocompatible artificial nerve Conduit. J. Cell. Mol. Med. 10 (3): pp 770-777.
2. Altschul, S. F., & Koonin, E. V. (1998). Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends in Biochemical Sciences*, *23*(11), 444-447.
3. Babb, P. L., Lahens, N. F., Correa-Garhwal, S. M., Nicholson, D. N., Kim, E. J., Hogenesch, J. B., … Voight, B. F. (2017). The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nature Genetics*, *49*(6), 895-903.
4. Blackledge, T. (2012). Spider silk: a brief review and prospectus on research linking biomechanics and ecology in draglines and orb webs. Journal of Arachnology. 40: pp. 1-12.
5. Blackledge, T., Kuntner, M., Agnarsson, I. (2011). The Form and Function of Spider Orb Webs: Evolution from Silk to Ecosystems. Academic Press. 41: pp. 175-262.
6. Blamires, S., Blackledge, T., Tso, I. (2017). Physicochemical Property Variation in Spider Silk: Ecology, Evolution, and Synthetic Production. Annu. Rev. Entomol. 62: pp. 443-460.
7. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421.
8. Clarke, T., Garb, J., Haney, R., Chaw, C., Hayashi, C., Ayoub, N. (2017). Evolutionary shifts in gene expresión decoupled from gene duplication across functionally distinct spider silk glands. Scientific Reports. 7:8393.
9. Clarke, T., Garb, J., Hayashi, C., Haney, R., Lancaster, A., Corbett, S., Ayoub, N. (2014). Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and functional processes of the Silk gland gene toolkit. BMC Genomics. 15:365.
10. Coddington, J. A. (1998). Spinneret silk spigot morphology : evidence for the monophyly of orb weaving spiders, Cyrtophorinae (Araneidae), and the group Theridiidae plus nesticidae. *The Journal of Arachnology*, 25.
11. Correa-Garhwal, S. M., Chaw, R. C., Clarke, T. H., Ayoub, N. A., & Hayashi, C. Y. (2017). Silk gene expression of theridiid spiders: implications for male-specific silk use. *Zoology*, *122*, 107-114.
12. Craig, C. & Riekel, C. (2002). Comparative architecture of silks, fibrous proteins and their encoding genes in insects and spiders. Comparative Biochemistry and Physiology Part B. 133: pp. 493-507.
13. Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., & Hormiga, G. (2018). Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. *Current Biology*, *28*(9), 1489-1497.e5.
14. Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., & Hormiga, G. (2018). Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. *Current Biology*, *28*(9), 1489-1497.e5.

15. Garrels, M. (2010). *Bash Guide for Beginners*. 173.
16. Giardine, B. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research*, *15*(10), 1451-1455.
17. Guerette, P. A., Ginzinger, D. G., Weber, B. H. F., & Gosline, J. M. (1996). Silk Properties Determined by Gland-Specific Expression of a Spider Fibroin Gene Family. *Science*, *272*(5258), 112-115.
18. Guinea, G., Elices, M., Plaza, G., Perea, G., Daza, R., Riekel, C., Agulló-Rueda, F., Hayashi, C., Zhao, Y., Pérez-Rigueiro, J. (2012). Minor Ampullate Silks from Nephila and Argiope Spiders: Tensile Properties and Microstructural Characterization. Biomacromolecules. 13: pp. 2087-2098.
19. Hayashi, C. & Lewis, R. (2001). Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. BioEssays. 23: pp.750-756.
20. Hayashi, C., Shipley, N., Lewis, R. (1999). Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. International Journal of Biological Macromolecules. 24: pp. 271-275.
21. Hinman, M. & Lewis, R. (1992). Isolation of a Clone Encoding a Second Dragline Silk Fibroin. The Journal of Biological Chemistry. 267 (27): pp. 19320-19324.
22. Kerfeld, C. A., & Scott, K. M. (2011). Using BLAST to Teach “E-value-tionary” Concepts. *PLoS Biology*, *9*(2), e1001014.
23. Kerfeld, C. A., & Scott, K. M. (2011). Using BLAST to Teach “E-value-tionary” Concepts. *PLoS Biology*, *9*(2), e1001014.
24. Lewis, R. (2005). Spider Silk: Ancient Ideas for New Biomaterials. Chem. Rev. 2006 (106): pp. 3762-3774.
25. Lewis, R. (2006). Spider Silk Production. Bionanotechnology. Pp. 61-78
26. Lombardi, S. & Kaplan, D. (1990). The Amino Acid Composition of Major Ampullate Gland Silk (Dragline) of Nephila clavipes (Araneae, Tetragnathidae). Journal of Arachnology. 18 (3): pp. 297-306.
27. Moon, M.-J., & An, J.-S. (2006). Microstructure of the silk apparatus of the comb-footed spider, Achaearanea tepidariorum (Araneae: Theridiidae): Silk apparatus of the comb-footed spider. *Entomological Research*, *36*(1), 56-63.
28. Sanggaard, K. W., Bechsgaard, J. S., Fang, X., Duan, J., Dyrlund, T. F., Gupta, V., … Wang, J. (2014). Spider genomes provide insight into composition and evolution of venom and silk. *Nature Communications*, *5*(1).
29. Singha, K., Maity, S., Singha, M. (2012). Spinning and Applications of Spider Silk. Frontiers in Science. 2 (5): pp. 92-100.
30. Tokareva, O., Michalczechen-Lacerda, V. A., Rech, E. L., & Kaplan, D. L. (2013). Recombinant DNA production of spider silk proteins: Recombinant spider silks. *Microbial Biotechnology*, *6*(6), 651-663.
31. Vollrath, F., & Knight, D. P. (2001). Liquid crystalline spinning of spider silk. *Nature*, *410*(6828), 541-548.

32. Wheeler, W. C., Coddington, J. A., Crowley, L. M., Dimitrov, D., Goloboff, P. A., Griswold, C. E., … Zhang, J. (2017). The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics*, *33*(6), 574-616.
33. Widhe, M., Johansson, J., Hedhammas, M., Rising, A. (2011). Current progress and limitations of spider silk for biomedical applications. Wiley Online Library. 97 (6): pp. 468-478.
34. Xu, M. & Lewis, R. (1990). Structure of a protein superfiber: Spider dragline silk. Proc. Natl. Acad. Sci. 87: pp. 7120-7124.
35. Xu, S., Xu, Z., Starrett, J., Hayashi, C., Wang, X. (2014). Cross-plane termal transport in micrometer-thick spider silk films. Polymer. pp. 1-9.

Pontificia Universidad JAVERIANA Bogotá

FACULTAD DE CIENCIAS
Anexo No. 6

Doctor
JORGE JÁCOME REYES
Director Carrera de Biología
Facultad de Ciencias

Respetado Doctor:

Con la presente comunicación, hacemos constar que el trabajo de grado titulado: Silk Specific Transcripts Across the Spider Phylogeny
realizado por los(as) estudiantes Christian C. Ramírez M. y _____,
ha sido revisado y corregido de acuerdo con las observaciones sugeridas por los jurados en la sustentación.

En constancia se firma, a los 6 días del mes de Junio del año 2019.

Cordialmente,

NOMBRE DIMITRI FORERO
FIRMA _____
DIRECTOR TRABAJO DE GRADO

NOMBRE Jeannette Gerwe
FIRMA Jeannette Gerwe
JURADO TRABAJO DE GRADO

NOMBRE Jorge Jácome
FIRMA _____
DIRECTOR CARRERA