DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Revealing Ancestral Arachnid Genes by Tracking Acetylcholinesterase Evolution in Spiders

Citation

Newman, Jordan Kian Mozaffarian. 2023. Revealing Ancestral Arachnid Genes by Tracking Acetylcholinesterase Evolution in Spiders. Master's thesis, Harvard University Division of Continuing Education.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37377280

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Revealing Ancestral Arachnid Genes by Tracking Acetylcholinesterase Evolution in Spiders

Jordan Kian Mozaffarian Newman

A Thesis in the Field of Biology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

November 2023

Copyright 2023 Jordan Kian Mozaffarian Newman

Abstract

Acetylcholinesterase (AChE) is a vital enzyme which plays a key role in regulation of neurotransmission in spiders and other animals. Because it is of critical importance, this protein is targeted by many toxins, including naturally occurring toxins and man-made pesticides. Due to the evolutionary pressures to circumvent the toxins found in their environment, the AChE genes (*ace*) of spiders have evolved over time. However, the phylogenetics of *ace* in spiders have not been extensively documented and have only been examined in detail in a single species, *Pardosa pseudoannulata*. Studying the evolutionary history of *ace* genes in spiders can help to determine how the respective roles and the importance of the different spider AChE proteins encoded by these genes have evolved over millions of years. Spiders are abundant predators which are important in controlling terrestrial arthropod populations worldwide, but they are under threat of being poisoned by the application of pesticides. Therefore, investigating the evolution of the various types of AChEs in spiders is also important for understanding the potential impact of pesticides on spiders. To address this knowledge gap, the relationships between previously published AChE amino acid sequences from 15 species of spiders were analyzed in this study. These analyses revealed a diverse array of AChE protein homologs and showed that multiple distinct *ace* genes often coexist in a single spider genome. Of note, a form of AChE with an amino acid sequence which is evolutionarily conserved even among distantly related spiders was also identified in this study. Since proteins with conserved sequences tend to be highly expressed, this finding suggests that

this form of AChE is expressed at a greater level than are other spider AChE proteins. This conserved spider AChE amino acid sequence is similar to AChE sequences not only from other arachnids, but also to those from animals as distantly related as insects, nematodes, and vertebrates. This study has also identified a separate clade of AChE with a less-conserved amino acid sequence, found only in spiders and in other arachnids. To my knowledge, this is the first evolutionary study of AChE across spider lineages, and these results should provide a basis for further research to facilitate protection of these indispensable species.

Acknowledgments

I would like to give my sincerest thanks to my thesis director, Dr. Gonzalo Giribet, who provided critical insights and guidance with this project. Additionally, I would like to thank my research advisor, Dr. James Morris, and my thesis template coordinator, Trudi Goldberg Pires. I would also like to thank my family for the help and support they have always given me.

Table of Contents

List of Tables

List of Figures

Chapter I.

Introduction

Considerable research has been devoted to spiders and to acetylcholinesterase protein in the past. However, these two topics have only rarely been investigated in combination, despite the fact that acetylcholinesterase is a vital component of spider physiology and spiders are a vital component of the biosphere.

Spider Ecology and Systematics

Spiders (Araneae) are a diverse, abundant, and broadly distributed group of terrestrial predators. They have important roles in both natural ecosystems and in agricultural food webs. Insects (Insecta) are among the most important prey for spiders. Since many important agricultural pests and disease vectors are insects, the presence of spiders is valuable for human societies as a mechanism for controlling the populations of insect pests. In agricultural areas, the density of the spider population strongly correlates with crop performance, a combination of the crop quality and crop yield. This is because predation by spiders reduces the population of insect pests and may also alter the insects' behavior by causing the insects to feed less (Michalko et al., 2019).

Phylogenetic Relationships of Spiders

To understand the genetic relationships among spiders, it is important to consider their evolutionary context among other closely related animals. Spiders, along with scorpions (Scorpiones) and other related groups, are members of the clade

Arachnopulmonata, which in turn is part of the broader arachnid clade (Arachnida) (Ontano et al, 2021). Other well-known representatives of Arachnida include mites (Acariformes) and ticks (Parasitiformes). Horseshoe crabs (Xiphosura) are traditionally considered the closest living relatives of arachnids (Lozano-Fernandez et al., 2019), though recent studies have identified horseshoe crabs as arachnids themselves (Ballesteros & Sharma, 2019; Ballesteros et al., 2022). Sea spiders (Pycnogonida) are relatives of arachnids and horseshoe crabs within the clade Chelicerata (Lozano-Fernandez et al., 2019). Together, Chelicerata, insects, and many other groups make up the arthropod clade (Arthropoda) (Figure 1A).

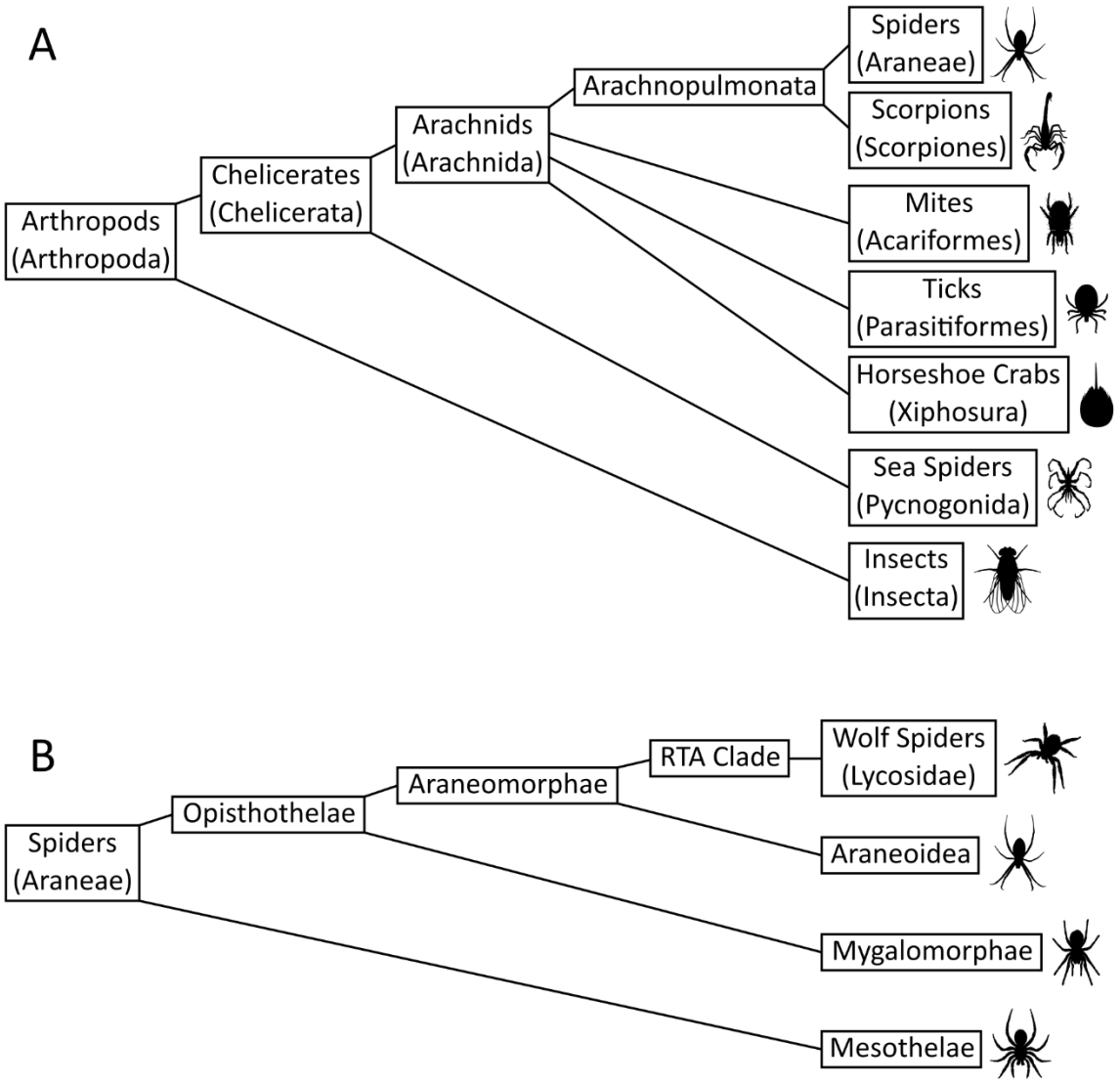

Figure 1. Phylogenetic Tree of Selected Arthropod Clades.

This figure shows the evolutionary relationships among clades that are relevant to this study. A. Relationships of spiders to other select arthropod clades. In accordance with recent studies (e.g., Ballesteros & Sharma, 2019; Ballesteros et al., 2022), horseshoe crabs are represented here as arachnids. B. Relationships among select spider clades, according to the results of Kulkarni et al. (2020).

The organization of clades within Araneae is also an area of ongoing study, with some relationships still under debate. What is not controversial is that spiders can be divided into the clades Mesothelae and Opisthothelae, and that Opisthothelae in turn can

be divided into Mygalomorphae and Araneomorphae (Fernández et al., 2018). The subordinate clades within Araneomorphae are less firmly established, but recent analyses have found Araneoidea and the RTA clade to be two major clades within Araneomorphae (Figure 1B). The RTA clade contains the lycosids, or wolf spiders (Lycosidae) (Kulkarni et al., 2020). Lycosids are relevant to this current study because one lycosid species, *Pardosa pseudoannulata*, is unique among spiders in that AChE has been studied in detail in this species (e.g., Lin et al., 2022).

Genetic Evolution

Proteins are described as evolutionarily conserved if their amino acid sequence has changed little over the course of evolution. Evolutionarily conserved proteins can be detected by comparing amino acid sequences from distantly related organisms. If the similarity is great enough, these sequences are more likely to be the result of evolutionary conservation than to be the result of convergent evolution (Fitch, W. M., 1970). Evolutionary conservation of protein-coding genes typically correlates with high protein expression levels (Drummond et al., 2005). One reason for this is because proteins with high abundance in an organism have a greater potential to harm the organism if they are misfolded into detrimental structures or if they form maladaptive aggregates. For highly expressed proteins, the great amount of protein material present means that even a small proportion of misfolded protein could potentially harm the organism, so these proteins must be constrained to sequences that minimize the possibility of misfolding (Drummond et al., 2005). Other contributing reasons for the greater conservation of highly expressed proteins include their greater risk of maladaptive interactions, their greater time and energy costs to produce, and their likelihood of being crucial for normal physiology.

Therefore, a protein-coding gene found to be evolutionarily conserved can be predicted to be highly expressed (Drummond et al., 2005). However, this correlation is complicated by the existence of chaperone complexes, which help proteins to fold properly. With chaperone complexes aiding the proper formation of proteins, abundant proteins can potentially evolve at a faster rate than they could without the chaperones, acquiring changes to their sequence without as much risk to the organism. Chaperone complexes can create exceptions to the trend that highly expressed proteins are evolutionarily conserved, but this trend is still observed in most cases (Agozzino & Dill, 2018).

Gene duplication is a key method by which organisms' genomes evolve and gain new genes. Without duplication, the probability that a gene would evolve a new function without interfering with its existing functions is unfavorable. However, a duplicated copy of a gene can lose or diminish its original functions and gain a new function with less risk to the organism, since the original copy of the gene can continue to perform its required functions. Gene duplication therefore creates opportunities for new gene functions to evolve in genomes. This promotes the adaptation of genomes to new conditions, which in turn promotes the diversification of organisms. As a result, duplicated genes are important genomic features.

There are several methods by which genes can be duplicated, including non-homologous chromosomal pairing during meiosis, reverse transcription of RNA into the genome, chromosome duplication, and genome duplication (Magadum et al., 2013). Genome duplications are particularly impactful for evolution, since they result in an additional copy of every gene in the genome. Genetic evidence suggests that approximately 450 million years ago, a whole-genome duplication occurred in

Arachnopulmonata, the lineage of Arachnida which would eventually produce spiders and scorpions, among other groups (Ontano et al, 2021). The genes duplicated by this event may have contributed greatly to the success and diversification of spiders and scorpions by helping them more quickly evolve to fill new niches (Schwager et al., 2017; Sharma et al., 2014).

Genes which originated from the same ancestral gene may be termed paralogs or orthologs. Paralogs are separate genes in the same organism that originated from duplication of an ancestral gene in that organism or its ancestor. Orthologs are homologous genes in different organisms that originated from a gene possessed by the common ancestor of those organisms. Therefore, orthologs are created by the divergence of a population into multiple species, not by gene duplication. Some of the species examined in this study are known to have multiple paralogs of *ace* (the gene for acetylcholinesterase) (Meng et al., 2017), and orthologs of *ace* were identified and compared across multiple species by evaluating their published amino acid sequences.

Acetylcholinesterase

Acetylcholinesterase (AChE) is a protein produced by a variety of organisms, including vertebrates (Vertebrata) and arthropods. It is a serine hydrolase enzyme with multiple functions, the best-known of which is hydrolysis of the neurotransmitter acetylcholine. This vital enzymatic function degrades excess acetylcholine, thus terminating the neuronal signal. Without AChE, these neuronal signals would continue to remain active even once no longer appropriate, which could result in paralysis or death of the organism (Soreq & Seidman, 2001). Measurements of AChE activity in spider nervous systems substantiate that this function is active in spiders (Meyer & Idel, 1977,

Meyer & Pospiech, 1977). Other processes that AChE has been reported to be involved with in vertebrates include neuronal growth, cell adhesion, synapse formation, activation of dopaminergic neurons, formation of amyloid protein fibers, and blood cell formation (Soreq & Seidman, 2001). In insects, AChE gene knockdown causes mortality of some individuals, malformation, reduced fecundity, and reduced growth, indicating that AChE plays developmental roles in insects as well (Meng et al., 2017).

Complete inhibition of AChE can be lethal in animals, as its absence severely disrupts neuronal signal transmission, among other physiological functions. Given the importance of AChE for survival, it is noteworthy that different animal species have different numbers of AChE gene paralogs. Insects have one to two, and nematodes (Nematoda) can have four (Meng et al., 2017). Jawed vertebrates (Gnathostomata) typically have two AChE gene paralogs, counting butyrylcholinesterase, an enzyme believed to have originated from a duplication of the AChE gene, while the two groups of jawless fish, hagfish and lampreys (Myxini and Petromyzontiformes) have just one AChE gene (Pezzementi et al., 2011).

The AChE protein is in the esterase family, a group of enzymes that hydrolyze ester molecules. Various toxins, including some insecticides, are among the molecules that esterases target and hydrolyze. Thus, esterases play a significant role in surviving toxin exposure (Montella et al., 2012). To correctly identify AChE sequences, it is important to recognize other similar proteins that share features in common with AChE. For example, other esterases have similarities to AChE in their structure and amino acid sequences, so with limited sequence data they could be confused with one another (Lockridge et al., 2018). Carboxylesterase is another important type of esterase. Different

carboxylases can vary greatly in their substrate specificities and can be encoded by very different gene sequences, though they retain some fundamental structural traits (Montella et al., 2012). The sequences of neurolignins also resemble those of AChEs. They are cell adhesion proteins and not esterases, but one of their domains, the α/β-hydrolase fold, is also present in AChEs, so neurolignins and AChEs share some similarities in their amino acid sequences (Leone et al., 2010).

AChEs in Insects and Spiders

Many of the best-studied arthropods are insects, so information on the AChEs of insects is important for making predictions about the AChEs of spiders, which have not been studied in as much detail as those of some insects. Most insects have two *ace* genes, which they inherited as a result of a gene duplication predating the origin of their clade (Kim & Lee, 2013). In insects with two *ace* genes (*ace1* and *ace2*), AChE1 is often the more important protein form, as it is found in greater abundance than AChE2 and is responsible for more acetylcholine catalysis. In some insects the transcription level of *ace1* is up to 250 times that of the transcription level of *ace2* (Kim & Lee, 2013), and in many insects, silencing of *ace1* has a much greater effect on their development and physiology than does silencing of *ace2* (Meng et al., 2017). AChE1 is also more evolutionarily conserved, having an amino acid sequence that is more similar to both vertebrate AChE and to nematode AChE1 than it is to insect AChE2 (Meng et al., 2016). However, in some insects the relative importance of AChE1 and AChE2 are reversed, and AChE2 is the form with greater catalytic importance (Meng et al., 2014). A study showed this to be the case in 33 species out of 100 tested, and notably, some insects, including *Drosophila melanogaster*, have lost *ace1* entirely (Kim & Lee, 2013).

Numerous predicted *ace* genes have been reported in spiders. Over 100 genes that have been predicted to code for AChE or AChE-like proteins, across 39 spider species, have been deposited into the National Library of Medicine genetic database as of 2023. However, most of these genes have not been studied beyond their initial identification (URL https://www.ncbi.nlm.nih.gov/). The number of *ace* paralogs in spider genomes is not well-studied, with the exception of the lycosid spider *Pardosa pseudoannulata*. This species has at least five *ace* paralogs, more than have been reported in any other organism so far. The total number of *ace* genes of *P. pseudoannulata* may be 17 or greater, as 12 other potential *ace* paralogs have been identified in the species' nuclear genome (Meng et al., 2015). The significance of the exceptional number of *ace* genes in this spider species is not fully understood.

Spiders have an AChE that is similar in its amino acid sequence to insect AChE1. One of the five known AChE proteins in *Pardosa pseudoannulata*, an AChE called PpAChE1, is most similar to insect AChE1, whereas the other four are not as similar to insect AChEs and form a separate clade along with other arachnid AChE amino acid sequences (Meng et al., 2017). Although the five *P. pseudoannulata* AChE proteins all exhibit the characteristic hydrolytic activity of AChE, they differ in several of their properties, such as optimal pH, concentration sensitivity, reaction speed, and sensitivity to various AChE inhibitors (Meng et al., 2017). Further research is necessary to understand the importance of each of these AChE proteins for the animal's survival.

Studying the evolution of AChE proteins in spiders could also help reveal how different spider groups have evolved in response to natural toxins, including abiotic toxins, those of their predators, and those of their prey. Several natural and synthetic

toxins act through the mechanism of AChE inhibition, so these proteins are important factors in toxin resistance.

The use of pesticides by humans can threaten ecosystems by poisoning ecologically important species, such as spiders. Spiders are natural predators which control terrestrial arthropod populations, but they are susceptible to pesticides that are used to control unwanted insects. The unintended killing of beneficial spiders could lead to reduced predation by spiders and could result in loss of ecological balance and overpopulation of other arthropods (Pékar, 2012).

Two major types of pesticides commonly used to control insects, carbamates and organophosphate pesticides, act via AChE inhibition (Engdahl, 2017). In insects, AChE mutations which confer pesticide resistance generally occur in *ace1*, suggesting that ensuring the functionality of AChE1 protein is important for the survival of the organism (Meng et al., 2016). Furthermore, silencing the *ace1* gene in insects causes a much greater increase in pesticide susceptibility than silencing the *ace2* gene (Meng et al., 2017).

There is typically only one copy of the *ace* gene in the mite *Tetranychus urticae*. Certain AChE mutations have been shown to confer pesticide resistance in this species (Kwon et al., 2010). However, some *T. urticae* individuals possess a different mechanism of resistance to monocrotophos pesticide in the form of duplicated copies of the *ace* gene. These mites express AChE in proportion to the number of *ace* copies they possess, which suggests that increased expression of AChE can be induced via gene duplication and that higher AChE protein levels help protect against AChE-targeting toxins (Kwon et al., 2010).

Pesticide resistance via gene mutation and gene duplication also occurs in a species of mosquito, *Culex pipiens*. In this species, pesticide exposure has selected for individuals with a mutant *ace1* allele with greater resistance to pesticides. However, this *ace1* codes for a protein with lower catalytic activity than the wild-type AChE1 protein. Although more resistant to toxins, this mutant *ace1* has a detrimental effect on the organism, as individuals with this allele accumulate greater levels of acetylcholine in their nervous systems. However, some of these mosquitos also carry an *ace1* gene duplication. The duplicated gene codes for an AChE1 that is pesticide-sensitive, but is catalytically more robust than the mutant form, restoring AChE protein activity to a more normal level and offsetting the disadvantage of the pesticide-resistant allele. Such gene duplications have occurred independently in multiple mosquito populations globally within the past 55 years, likely as adaptations to the use of pesticides by humans (Labbé et al., 2007).

Of interest with respect to the effects of pesticides on spiders, the AChE proteins of *Pardosa pseudoannulata* (designated as PpAChE) are known to be differentially inhibited by certain types of pesticides. PpAChE2 is less affected by carbamates than are the other AChEs, while PpAChE3 is more affected by organophosphates (Meng et al., 2016). Additionally, the AChE proteins of this spider have been observed to function in a compensatory fashion. Three of its five AChEs, PpAChE1, PpAChE2, and PpAChE5, are normally highly expressed in the brain, but exposure to organophosphate pesticides further increases expression of PpAChE5 in the brain. PpAChE5 catalyzes acetylcholine more efficiently and is less sensitive to organophosphates than either PpAChE1 or PpAChE2. Therefore, the increase in PpAChE5 expression likely serves as a defense

mechanism, allowing the spider to maintain AChE activity in its brain when exposed to an AChE-inhibiting toxin (Lin et al., 2022). Studying the distribution of *ace* paralogs and orthologs in spiders may help predict how spider lineages are differentially affected by various pesticides and may lead to the development of less ecologically harmful regimens of pesticide application.

Research Question

In this study, I intend to discern the relationships between the AChE protein sequences found in spiders, and to use that information to infer how the different *ace* genes in spiders have originated and diversified. To do so, I have generated and analyzed phylogenetic trees from previously published AChE amino acid sequences of spiders and other animals. On the basis of the results of Meng et al. (2017), I hypothesize that spiders have multiple *ace* gene paralogs, and that one of these paralogs codes for a protein that is similar to insect AChE1 in its sequence and in its greater importance for the organism's survival.

Chapter II.

Materials and Methods

Over the course of this study, three phylogenetic trees were inferred based on AChE amino acid sequence data downloaded from the National Library of Medicine database (URL https://www.ncbi.nlm.nih.gov/). The sequences analyzed in this study are predicted amino acid sequences from previously published results, based on nucleotide sequences of putative protein-coding genes.

Tree One

The first stage of this study was conducted in order to identify evolutionary relationships among published spider AChE protein sequences. Spider amino acid sequences in the National Library of Medicine database that were labeled as AChE were downloaded for analysis. An amino acid alignment was generated for these sequences using Multiple Alignment with Fast Fourier Transform (MAFFT) (Madeira et al., 2022) with the following settings: BLOSUM62 matrix, 1.53 gap open penalty, 0.123 gap extension, 'tree rebuilding number' 2, 'guide tree output' on, 'maxiterate' 2, no fast Fourier transform (FFTS).

A phylogenetic tree, referred to in this study as Tree One, was inferred from this alignment using IQ-TREE (Nguyen et al., 2015; Trifinopoulos et al., 2016) using the following settings: automatic substitution model (Kalyaanamoorthy et al., 2017), no 'FreeRate heterogeneity,' ultrafast bootstrap analysis (Hoang et al., 2018), 1000 bootstrap

alignments, 1000 maximum iterations, minimum correlation coefficient 0.99, 1000 replicates of SH-aLRT branch test, perturbation strength 0.5, IQ-TREE stopping rule 100.

Tree Two

The second stage of this study was conducted in order to find additional spider AChE protein sequences that fall within a clade identified in Tree One which included the sequence PpAChE1, since this clade showed the greatest degree of evolutionary conservation. To find additional spider AChE sequences similar to that of PpAChE1 that were not already included in Tree One, an amino acid alignment was generated using MAFFT, with the same settings as above, aligning PpAChE1 with the amino acid sequences recovered as its closest relatives in Tree One.

Amino acid sequences similar to the aligned sequence with the greatest total identity to the other aligned sequences were sought using protein Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) using the following settings: standard databases, taxon Araneae, blastp algorithm, max target sequences 100, automatically adjust parameters for short input sequences, expect threshold 0.05, word size 5, max matches in a query range 0, BLOSUM62 matrix, gap existence cost 11, gap extension cost 1, conditional compositional score matrix adjustment.

To study the relationships among PpAChE1 and the other sequences found using BLAST, another amino acid alignment was generated using MAFFT, with the same settings as above. This alignment consisted of PpAChE1, the spider AChE sequences recovered as its closest relatives in Tree One, and the new sequences found using BLAST, as well as three other sequences from Tree One chosen arbitrarily to serve as outgroups to the clade containing PpAChE1.

A phylogenetic tree, referred to in this study as Tree Two, was inferred from this alignment using IQTREE, with the same settings as Tree One.

Tree Three

The third stage of this study was conducted in order to determine how the spider AChE clades identified in this study are related to the AChE proteins of other animals. To find AChE sequences from animals that are more closely related to spiders, amino acid sequences similar to the aligned spider AChE sequence with the greatest total identity to the other aligned sequences were sought using protein BLAST using the same settings and the same target sequence as for Tree Two, except that a separate search was performed for each of the following taxa: Pycnogonida, Xiphosura, and non-Araneae non-Acariformes non-Parasitiformes non-Xiphosura Arachnida. All resulting sequences with a greater likelihood (measured by E value) to match the target sequence than the likelihood of insect esterase to match an insect AChE sequence were downloaded.

An amino acid alignment was generated using MAFFT, with the same settings as for Tree One. This alignment consisted of PpAChE1, all other spider AChE sequences previously tested in this study, 11 non-spider AChE sequences, the non-spider sequences found using BLAST, and esterase from *Drosophila melanogaster*.

A phylogenetic tree, referred to in this study as Tree Three, was inferred from this alignment using IQTREE, with the same settings as for Tree One.

Graphics

Phylogenetic trees inferred in this study were converted to figures using FigTree v1.4.4 (tree.bio.ed.ac.uk/software/figtree), then edited using Inkscape 1.2.2 (The Inkscape

Project). Other figures were created using Inkscape 1.2.2. All animal images used in figures in this study are in the public domain.

Chapter III.

Results

Evolutionary events have created new *ace* genes over time. When an *ace* gene is duplicated, it becomes a pair of *ace* gene paralogs. As species diverge from each other, *ace* gene orthologs result from the *ace* genes of each of those new species. The AChE proteins encoded by these genes share some similarities as a result of their common ancestry, but similarities between AChE and other proteins can also result from convergent evolution. As originally shown by Fitch (1970), computational methods can indicate whether a similarity between amino acid sequences is due to a common ancestry between those sequences or merely due to convergent evolution. Modern digital computational methods such as those used in this study allow analyses of relationships between amino acid sequences to accurately analyze greater amounts of data at greater speeds (Nguyen et al., 2015; Trifinopoulos et al., 2016). The phylogenetic trees inferred in this study use the relationships between AChE amino acid sequences to conclude how paralogs and orthologs of *ace* genes in spiders and other animals may have originated, and which other genes they may be related to.

Spider Sequences Labeled as AChE in the National Library of Medicine Database

In a search conducted in February 2023, 76 spider amino acid sequences labeled as AChE were downloaded from the National Library of Medicine database. These sequences were originally obtained from nine separate spider species by the various

researchers who sequenced and submitted them. Sequences marked as being contaminated by foreign genetic sequences were not included in the results of this search.

Tree One and Identification of a Conserved Clade Within Spider AChE Sequences

A phylogenetic tree of spider amino acid sequences labeled as AChE in the National Library of Medicine database, Tree One, was constructed using IQ-TREE (Nguyen et al., 2015; Trifinopoulos et al., 2016). The IQ-TREE software found WAG+F+G4 to be the optimal amino acid evolution model for this tree. Ten of the sequences tested in this tree were found to be highly similar to each other despite originating from distantly related spiders (Figure 2, top right). These ten sequences comprise a clade with short branch lengths relative to the branch lengths among the other amino acid sequences. The clade has a bootstrap support value of 100%.

The similarity of these ten sequences suggests that they are encoded by a highly conserved paralog of *ace* in spiders. By contrast, the other AChE amino acid sequences are much more disparate to each other, as indicated by the longer branch lengths among them. This conserved category of AChE proteins, newly identified in this current study, is referred to as "Spider AChE1," since it includes PpAChE1 from *Pardosa pseudoannulata* (labeled as *Pardosa pseudoannulata* 1; Figure 2). Throughout this current study, Spider AChE1 is defined as the least inclusive clade containing all ten of these amino acid sequences, to ensure that this clade remains manageable in scope and does not expand unnecessarily to encompass too many additional sequences.

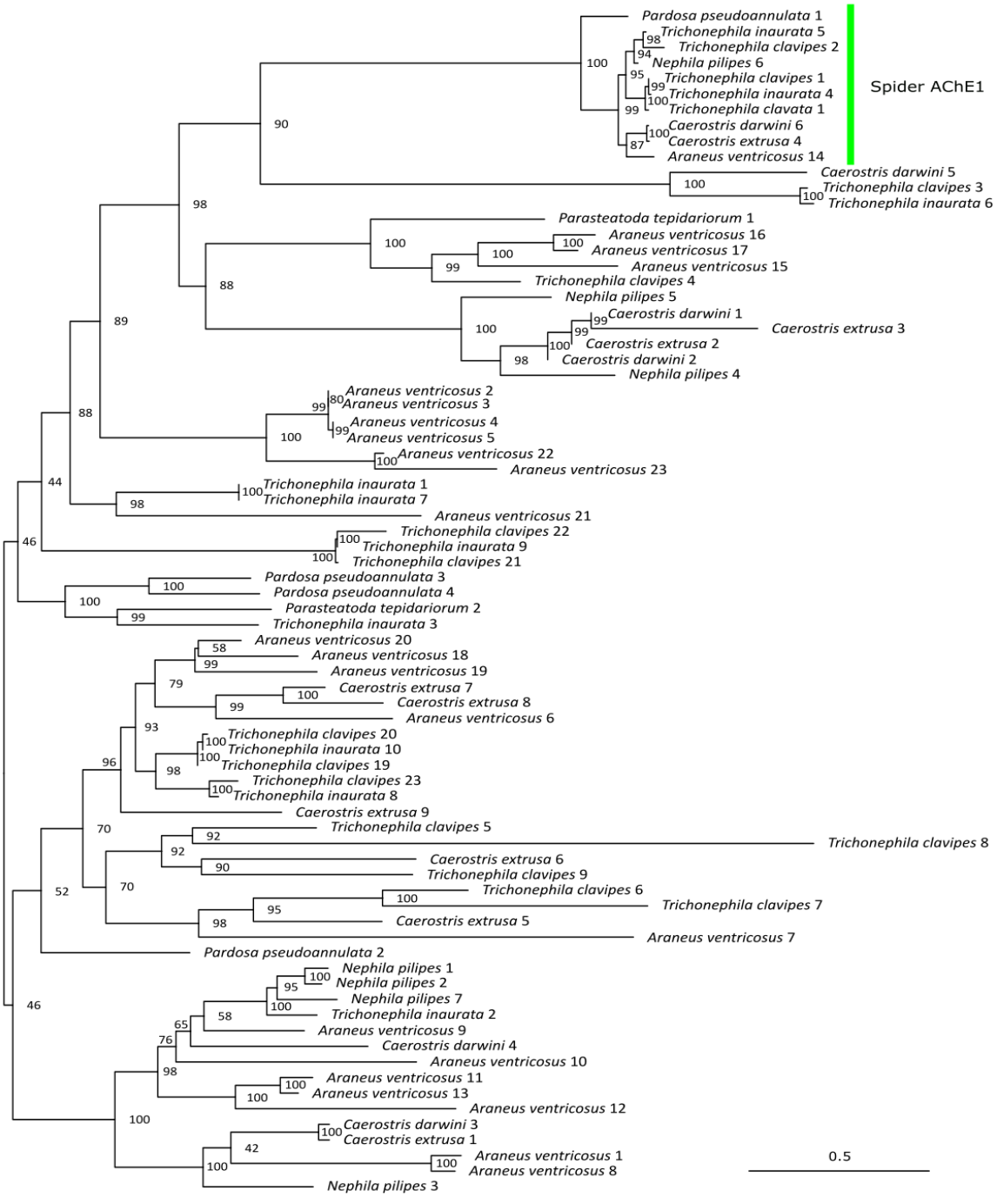

Figure 2. Tree One, Spider Amino Acid Sequences Labeled as AChE in the National Library of Medicine database.

Horizontal distance represents branch lengths. Numbers at nodes indicate bootstrap support values. The clade labeled Spider AChE1 has a highly conserved amino acid sequence compared to other AChE sequences tested on this tree, despite containing sequences from more distantly related spiders.

Apart from *Trichonephila clavata*, all of the species with AChE amino acid sequences within the Spider AChE1 clade also had other AChE sequences which were recovered outside of the Spider AChE1 clade on the phylogenetic tree. This is particularly true of *Araneus ventricosus*. The 23 amino acid sequences from *A. ventricosus* included in Figure 2 may represent as many as 17 different AChE protein paralogs, judging by the branch lengths between these sequences, which are often comparable to branch lengths between different species' AChE sequences. This result is not unprecedented, considering the presence of 17 potential AChE protein paralogs in *Pardosa pseudoannulata* (Meng et al., 2015).

Tree Two and Identifying Additional Spider AChE Sequences

Alignment of the ten identified Spider AChE1 amino acid sequences using MAFFT (Madeira et al., 2022) indicated that the *Nephila pilipes* 6 sequence (NCBI database accession BMAW01098500) had the greatest total percent identity to the other sequences in this alignment. Because of its similarity to the other Spider AChE1 sequences, *Nephila pilipes* 6 was used in this current study as a representative of the Spider AChE1 clade for the purpose of finding additional sequences similar to Spider AChE1.

To identify Spider AChE1 proteins from additional species, BLAST was used to search the National Library of Medicine database for spider amino acid sequences that are more similar to the Spider AChE1 sequence *Nephila pilipes* 6 than they are to *Pardosa pseudoannulata* AChE4 (PpAChE4), as measured by E value. Twelve such sequences that had not already been included in Tree One were found.

A new phylogenetic tree, Tree Two, was then constructed using IQ-TREE from the ten already-identified Spider AChE1 amino acid sequences, three other spider AChE sequences chosen arbitrarily to serve as outgroups of Spider AChE1, and the 12 additional spider amino acid sequences found using BLAST. The IQ-TREE software found LG+I+G4 to be the optimal amino acid evolution model for this tree. This comparison revealed that some of the additional spider ACHE amino acid sequences nested within the Spider AChE1 clade (Figure 3). Interestingly, one of the sequences found using BLAST, *Lampona murina* 1 (NCBI database accession ON226730), was recovered as the sister group to the existing Spider AChE1 clade, with a bootstrap support value of 100%.

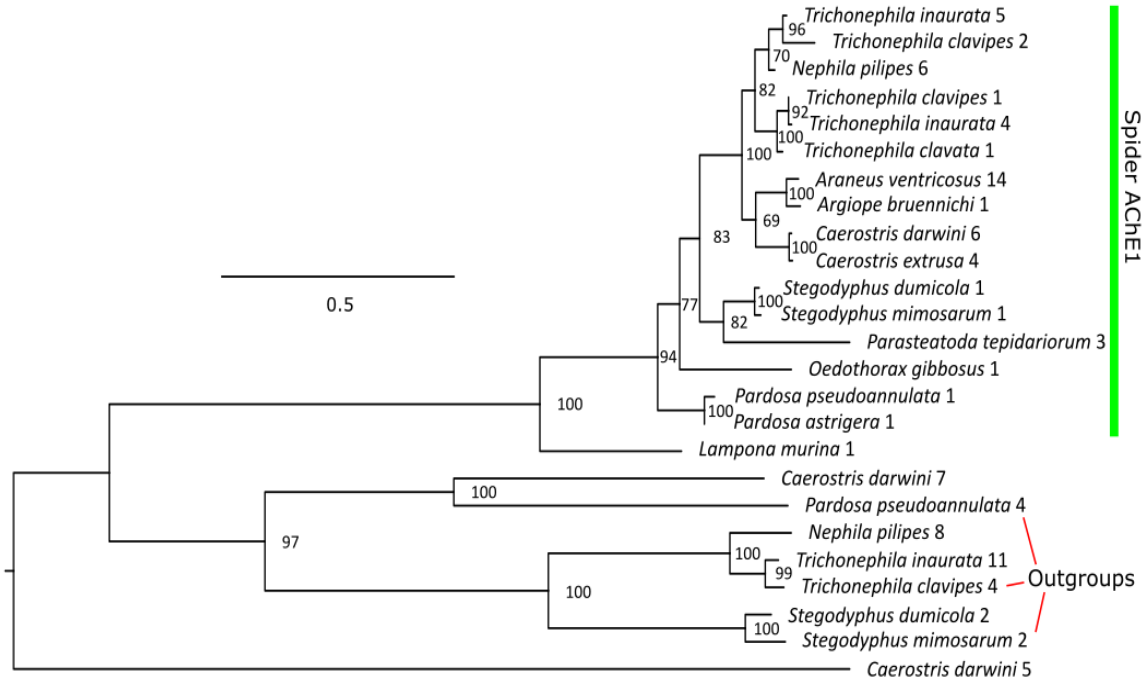

Figure 3. Tree Two, Spider AChE Amino Acid Sequences.

Horizontal distance represents branch lengths. Numbers at nodes indicate bootstrap support values.

Tree Three and Non-Spider Orthologs of Spider AChE1

To determine the relatedness of proteins in non-spider animals to spider AChE, AChE amino acid sequences from various animals were downloaded from the National Library of Medicine database. These sequences consisted of vertebrate AChE (represented by *Tetronarce californica*), the four nematode AChE proteins (*Caenorhabditis elegans*), AChEs from an insect with two AChE proteins (*Blattella germanica*), a mite AChE (*Tetranychus urticae*), and AChEs from a tick with three AChE proteins (*Rhipicephalus microplus*).

To find AChE sequences from additional relatives of spiders, BLAST was used to search the National Library of Medicine database for amino acid sequences similar to Spider AChE1 from scorpions, horseshoe crabs, and sea spiders. The sequences found via BLAST consisted of 50 sequences from the scorpion Centruroides sculpturatus, 45 from the horseshoe crab Limulus polyphemus, and 42 from the sea spider Nymphon striatum.

A new phylogenetic tree, Tree Three, was constructed using IQ-TREE from the 16 Spider AChE1 sequences, the Spider AChE1-like sequence *Lampona murina* 1, the three other spider AChE sequences previously chosen in this study as outgroups, 11 non-spider AChE amino acid sequences, and the 137 non-spider amino acid sequences which were matched with Spider AChE1 via BLAST. The IQ-TREE software found VT+I+G4 to be the optimal amino acid evolution model for this tree.

Tree Three recovered Spider AChE1 within a clade also containing insect AChE1, tick AChE1, mite AChE, and scorpion and horseshoe crab amino acid sequences (Figure 4). Nematode AChE1 and vertebrate AChE were recovered within this clade as well. Meanwhile, most of the other spider AChE sequences were recovered within a separate

clade, which also contained the tick AChE2 and AChE3 sequences and scorpion and horseshoe crab AChE sequences. This result indicates that some spider AChE proteins are more similar in amino acid sequence to vertebrate AChEs than they are to other spider AChEs, despite the very distant evolutionary relationship between spiders and vertebrates. Such a great degree of similarity between amino acid sequences from distantly related animals implies that vertebrate AChE and Spider AChE1 are highly conserved proteins with a common ancestry.

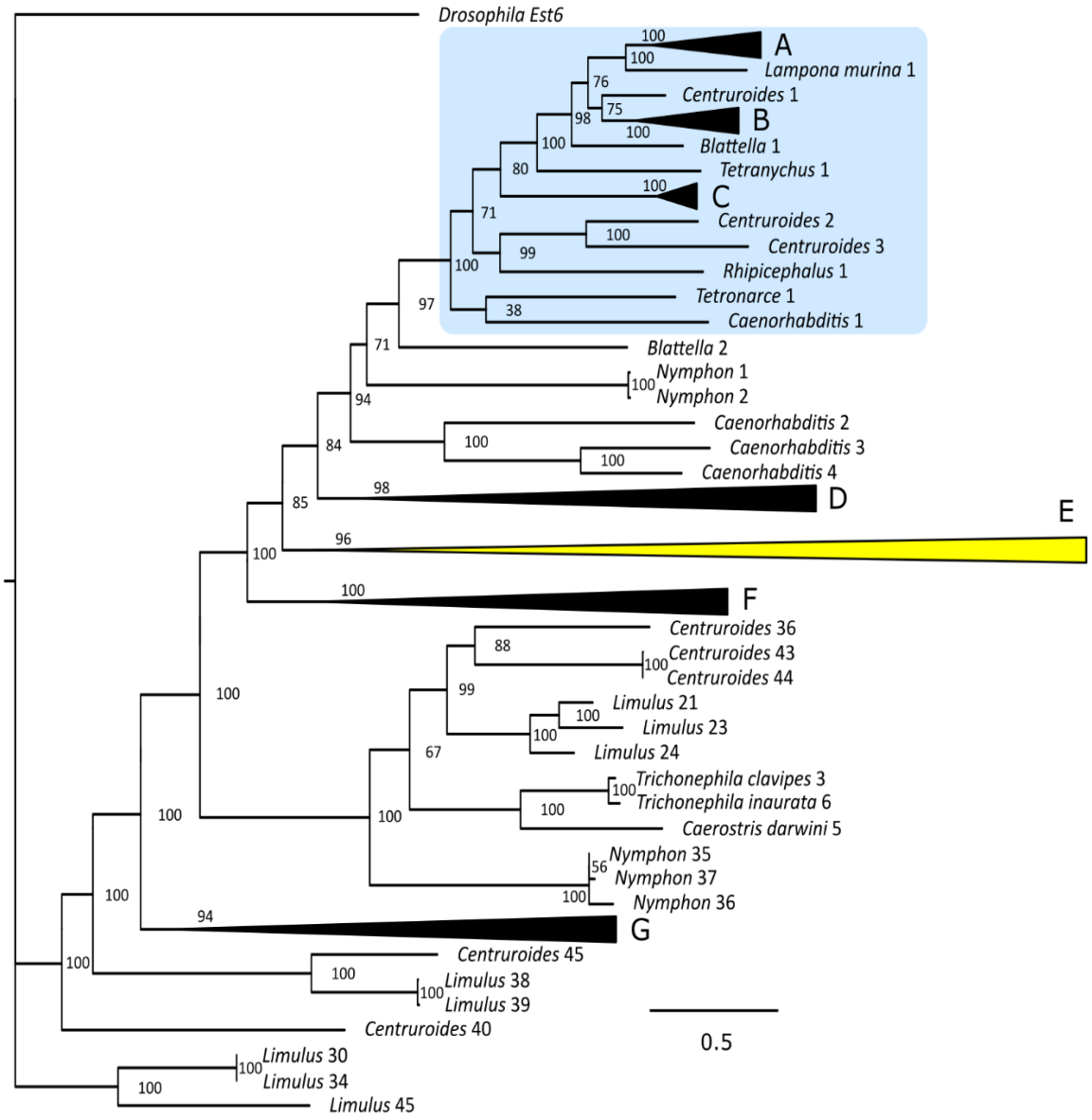

Figure 4. Tree Three, Amino Acid Sequences of AChE and Similar Gene Products.

Horizontal distance represents branch lengths. Numbers at nodes indicate bootstrap support values. Drosophila Est6 *is* Drosophila *esterase. Certain clades are represented as triangles for increased clarity of the figure. Clade A is Spider AChE1, containing 16 Araneae sequences. Clade B contains seven* Limulus *sequences. Clade C contains four* Limulus *sequences. Clade D contains 32* Nymphon *sequences. Clade E contains the Arachnid Non-AChE1 clade, shown in detail in Figure 5. Clade G contains five* Nymphon *sequences. Clade H contains 29* Centruroides *and* Limulus *neurolignin-like sequences (Battelle et al., 2016; Schwager et al., 2017). The shaded box indicates the AChE1-Like clade and unites Spider AChE1 with insect AChE1 and other related AChEs.*

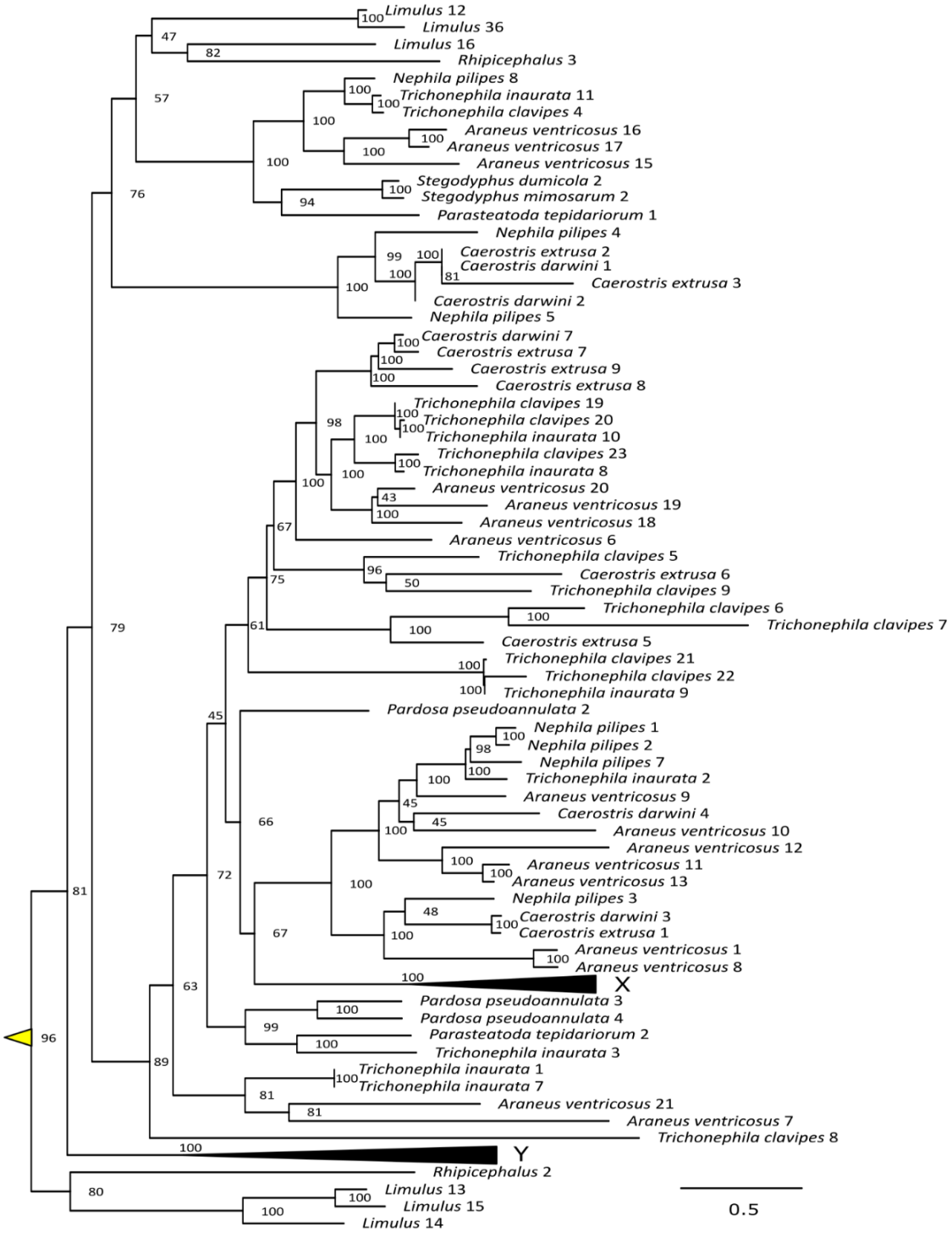

Figure 5. Tree Three Continued, Arachnid Non-AChE1 Amino Acid Sequences.

Horizontal distance represents branch lengths. Numbers at nodes indicate bootstrap support values. Certain clades are represented as black triangles for increased clarity of the figure. Clade X contains six Araneus ventricosus *sequences. Clade Y contains 33* Centruroides *sequences.*

Chapter IV.

Discussion

The results of this study show a diverse complement of AChE proteins, suggesting the existence of a diverse set of *ace* gene paralogs in spiders (Figure 2). Of these, Spider AChE1, a protein clade identified in this study, remains substantially more conserved than other AChE protein clades, across spiders as distantly related as RTA-clade spiders (i.e., *Pardosa*) and araneoids. The highly conserved nature of Spider AChE1 is underscored by the temporal separation between these clades, which diverged from each other over 200 million years ago (Carlson & Hedin, 2017). Because high evolutionary conservation of a protein typically correlates with high expression level of that protein (Agozzino & Dill, 2018), Spider AChE1 is likely expressed at a high level across a wide range of spider species. This contention that Spider AChE1 is more highly expressed than the other spider AChE proteins should be tested further, since it is possible that chaperone complexes could be contributing to the greater evolution rate of non-AChE1 proteins, raising the possibility that they also are expressed at high levels (Agozzino & Dill, 2018).

By comparing the Spider AChE1 clade to the two insect AChE proteins, I found Spider AChE1 to be more similar to insect AChE1 than to insect AChE2. AChE1 has a more important catalytic role in insects than its counterpart, AChE2 (Meng et al., 2016), adding further support to the idea that Spider AChE1 is an important protein, and perhaps the most physiologically important AChE in spiders. Furthermore, Spider AChE1 is more similar to vertebrate AChE and to nematode AChE than it is to other spider AChE

proteins, which further demonstrates the great degree of evolutionary conservation of Spider AChE1.

Comparison of spider AChE amino acid sequences with non-spider AChE sequences in this study revealed that spider AChE proteins fall into multiple clades (Figures 4, 5). The first of these clades, which I will refer to as "AChE1-Like", includes Spider AChE1, an AChE from the venom glands of the spider *Lampona murina*, some scorpion AChEs, some horseshoe crab AChEs, insect AChE1, mite AChE, tick AChE1, vertebrate AChE, and nematode AChE1. The second clade, which I will refer to as "Arachnid Non-AChE1," includes a variety of other spider AChEs, other horseshoe crab AChEs, tick AChE2 and AChE3, and other scorpion AChEs. Both clades are well-supported, with 100% and 96% bootstrap support respectively.

All spider AChEs tested in this study fell into either the AChE1-Like or Arachnid Non-AChE1 clade, except for three: *Trichonephila clavipes* 3, *Trichonephila inaurata* 6, and *Caerostris darwini* 5, which formed a clade with scorpion and horseshoe crab sequences previously identified as carboxylesterases (Battelle et al., 2016; Schwager et al., 2017). This implies that these three spider amino acid sequences are carboxylesterases, which would contradict their original identification as AChE proteins (Kono, Nakamura, et al., 2021; Kono, Ohtoshi, et al., 2021). These sequences should be evaluated further to determine whether they are in fact carboxylesterases.

Notably, no AChE1-Like sequence was found from the sea spider *Nymphon striatum*. The 42 protein sequences from this species with the greatest similarity to Spider AChE1 were analyzed, and these were derived from data in a study which sequenced the full genome of this species (Jeong et al., 2020). Therefore, it is unlikely that an AChE1-

Like sequence exists in *N. striatum* and was absent from this current analysis, given the completeness of its published genome. To my knowledge, this is the initial report that the gene for the AChE1-Like protein is absent in sea spiders. Two possible explanations for this absence are that the ancestor of *N. striatum* lost its gene for the AChE1-Like protein at some point after its evolutionary divergence from arachnids, or that so many evolutionary changes have occurred in the sea spider's AChE1-Like sequence that it can no longer be identified as such. In turn, this has interesting implications for the evolution of the *ace* genes of *N. striatum*.

AChE is a necessary protein in animals, and animals as disparate as arthropods and vertebrates usually have AChE1-Like proteins. Flies in the clade Cyclorrhapha are an exception to this rule, as they lack the *ace1* gene. This represents an extreme version of a condition which occurs in some other insect groups, where AChE2 has a more important functional role in the organism than AChE1 (Kim & Lee, 2013). Similar evolutionary processes in both Cyclorrhapha and sea spiders, which are not closely related and diverged over 500 million years ago (Dohrmann & Wörheide, 2017), may have led to the replacement and loss of the gene for the AChE1-Like protein. To my knowledge there has not yet been an investigation of the properties of AChEs in sea spiders. Such a study could help determine how the other AChE proteins of *N. striatum* make up for the absence of the AChE1-Like enzyme.

Spiders, scorpions, ticks, and horseshoe crabs all have AChE proteins in both the AChE1-Like and Arachnid Non-AChE1 clades, implying that their common ancestor, an early arachnid or arachnid-ancestor, already had at least two *ace* genes, each coding for a protein in one of the two AChE clades identified in this study. A whole-genome

duplication occurred in the lineage which would produce the arachnid clade Arachnopulmonata, which includes spiders and scorpions but not ticks or horseshoe crabs (Ontano et al., 2021). However, this duplication event does not serve as an explanation for the original divergence of the AChE1-Like clade from the Arachnid Non-AChE1 clade, since ticks and horseshoe crabs have both forms of AChE despite not being members of Arachnopulmonata. The whole-genome duplication in Arachnopulmonata could have produced a total of four AChEs from an ancestor with two *ace* genes, but this duplication would still be insufficient to explain the number of *ace*paralogs which are found in a single spider species since this number can exceed four, as seen with the five documented paralogs in *Pardosa pseudoannulata* (Meng et al., 2017). Sequences of AChE proteins from additional arachnids could help fill in more details of the protein's evolutionary history in these animals.

The evolutionary placement of horseshoe crabs within the clade Chelicerata has been subject to some debate, with the traditional view being that horseshoe crabs are the sister clade to arachnids. Some recent studies instead support the placement of horseshoe crabs as derived arachnids related to Arachnopulmonata, a clade including spiders and scorpions (Ballesteros & Sharma, 2019; Ballesteros et al., 2022). The findings of this study are consistent with the latter interpretation, given that certain horseshoe crab AChE amino acid sequences (Figure 4, labeled as B) were recovered within a clade with 76% bootstrap support consisting of only horseshoe crab, spider, and scorpion AChE sequences. However, this finding should be treated only as tentative support for this relationship, since the results of phylogenetic analyses in this study do not correspond exactly to known phylogenetic relationships. For example, this study recovered Spider

AChE1 protein as more closely related to insect AChE1 than to mite AChE1 (Figure 4), despite the fact that spiders and mites are both members of the arachnid clade while insects are not.

The great number of spider AChE sequences within the Arachnid Non-AChE1 clade suggests that the abundance of AChE proteins in spiders may have resulted from duplication of genes coding for Arachnid Non-AChE1. Why duplications of Arachnid Non-AChE1 would be more frequent than duplications of Spider AChE1 is not obvious, but it may be that certain arachnids derived an adaptive benefit from having additional copies of the gene for Arachnid Non-AChE1. Mites of the species *Tetranychus urticae* have adapted to pesticides via duplications of their *ace* gene, which causes them to produce more AChE (Kwon et al., 2010). Perhaps at some point in their evolutionary history, spiders evolved a duplication of a gene for Arachnid Non-AChE1, and this improved their fitness by causing them to produce more AChE, as it has in *T. urticae*. The benefit of the additional *ace* copies in spiders might have been related to resistance to natural toxins, or might have been related to the other effects of AChE. Alternatively, it could be that duplications of the gene for Spider AChE1 are deleterious, perhaps by interfering with the normal functions of AChE1 or other proteins, or by harming the organism with detrimentally high levels of AChE1 enzymatic activity.

Thirty-three sequences from the scorpion *Centruroides sculpturatus* also fell into the Arachnid Non-AChE1 clade in this study, but these formed a single clade within the Arachnid Non-AChE1 clade (Figure 5, labeled as Y), suggesting that the proliferation of the genes for Arachnid Non-AChE1 proteins in scorpions was independent of the similar proliferation that occurred in spiders. Research into the physiological functions of

different AChE proteins in spiders and scorpions would be useful for determining the evolutionary cause of these proliferations of Arachnid Non-AChE1. It is worth noting that, unlike the other arachnids included in this study, spiders and scorpions are venomous terrestrial predators. Their related lifestyles could have driven convergent selection for the large number of duplicated genes coding for Arachnid Non-AChE1 proteins.

Although unlikely, there is the possibility that, contrary to the apparent results of phylogenetic analyses in this study, the Arachnid Non-AChE1 clade is not monophyletic, which would explain the long branch lengths within the apparent clade. If this is the case, then there may have been multiple duplication events which produced *ace* gene paralogs other than *ace1* in spiders, and *ace* gene paralogs newly duplicated from *ace1* may have undergone convergent evolution as their roles shifted. This scenario could account for the potentially false impression that Arachnid Non-AChE1 is a monophyletic group. A comparison of Arachnid Non-AChE1 sequences including sequences from additional species could determine with greater confidence whether this clade is truly monophyletic.

The Spider AChE1 amino acid sequence tree generated in this current study also shows that *Trichonephila clavipes* and *Trichonephila inaurata* each exhibit two AChE protein sequences that fall within the Spider AChE1 clade (Figure 2). This indicates a duplication of the gene for Spider AChE1 in the *Trichonephila* lineage that could have occurred prior to the divergence of these two species from each other, which took place approximately 20 million years ago (Turk et al., 2020). These *ace* paralogs have not yet become vastly different in the time since the duplication. Another possibility, which is not mutually exclusive with the above, is that the *ace* genes for these two proteins have

undergone subfunctionalization in a way that has constrained both to remain similar to the ancestral Spider AChE1 sequence. Such a duplication would be an exception to the apparent rule, as it seems that duplicated *ace* genes in spiders are more often duplicates of the genes for Arachnid Non-AChE1.

In conclusion, I have found in this study that AChE amino acid sequences in spiders and other arachnids can be categorized into two major clades. One of these clades, which I call AChE1-Like, is exemplified by a conserved protein sequence which is similar to that of AChE1 in insects and in nematodes, and also similar to that of vertebrate AChE. The degree of conservation of this protein sequence is striking considering that the most recent common ancestor of spiders and vertebrates, at the base of the clade Bilateria, is inferred to have existed approximately 700 million years ago (Dohrmann & Wörheide, 2017). The conserved sequence of this protein across disparate species suggests that it is expressed at high levels. The other clade, which I call Arachnid Non-AChE1, has a far less conserved protein sequence, and numerous paralogs of genes for Arachnid Non-AChE1 occur in at least some spider and scorpion species. The findings in this study shed light on the evolution and diversity of AChE proteins across spider and other arachnid lineages for the first time, which should aid future efforts to minimize the negative impact of pesticide use on the biosphere's numerous and important spider predators.

Appendix 1.

Sequences Used in This Study

The following sequences were accessed from the National Library of Medicine database and analyzed in this study.

Table 1. Spider AChE Amino Acid Sequences

NCBI Accession	Species	Assigned Number	Description
AHB20142.1	*Pardosa pseudoannulata*	1	Labeled AChE in NCBI database
ANQ45782.1	‘	2	‘
ANQ45783.1	‘	3	‘
ANQ45784.1	‘	4	‘
XP_015927912.2	*Parasteatoda tepidariorum*	1	‘
NW_024969791	‘	2	‘
XP_042903818.1	‘	3	Found via BLAST
GFU18170.1	*Nephila pilipes*	1	Labeled AChE in NCBI database
GFU18172.1	‘	2	‘
GFU18177.1	‘	3	‘
GFU18181.1	‘	4	‘
GFU18184.1	‘	5	‘
GFS85149.1	‘	6	‘
GFT23927.1	‘	7	‘
GFS89061.1	‘	8	Found via BLAST
GFQ99676.1	*Trichonephila clavata*	1	Labeled AChE in NCBI database
GFU73840.1	*Trichonephila clavipes*	1	‘
GFU73845.1	‘	2	‘
GFV21411.1	‘	3	‘
GFV35312.1	‘	4	‘
GFV80846.1	‘	5	‘
GFV80849.1	‘	6	‘
GFV80888.1	‘	7	‘

NCBI Accession	Species	Assigned Number	Description
GFV80890.1	‘	8	‘
GFV80897.1	‘	9	‘
GFV92734.1	‘	19	‘
GFV92738.1	‘	20	‘
GFV92751.1	‘	21	‘
GFV92758.1	‘	22	‘
GFV92770.1	‘	23	‘
GFS53526.1	*Trichonephila inaurata*	1	‘
GFY61515.1	‘	2	‘
GFY59257.1	‘	3	‘
GFY56267.1	‘	4	‘
GFY56268.1	‘	5	‘
GFY55080.1	‘	6	‘
GFY54282.1	‘	7	‘
GFY40062.1	‘	8	‘
GFY40063.1	‘	9	‘
GFY40064.1	‘	10	‘
GFY46079.1	‘	11	Found via BLAST
GBM91030.1	*Araneus ventricosus*	1	Labeled AChE in NCBI database
GBO10039.1	‘	2	‘
GBO10025.1	‘	3	‘
GBO09986.1	‘	4	‘
GBO09985.1	‘	5	‘
GBN79984.1	‘	6	‘
GBN69267.1	‘	7	‘
GBM77062.1	‘	8	‘
GBM77064.1	‘	9	‘
GBM77067.1	‘	10	‘
GBM77076.1	‘	11	‘
GBM77079.1	‘	12	‘
GBM77080.1	‘	13	‘
GBM73330.1	‘	14	‘
GBM53520.1	‘	15	‘
GBM53521.1	‘	16	‘
GBM53525.1	‘	17	‘
GBM49428.1	‘	18	‘
GBM49431.1	‘	19	‘
GBM49435.1	‘	20	‘
GBM49084.1	‘	21	‘
GBM49089.1	‘	22	‘

NCBI Accession	Species	Assigned Number	Description
GBM49093.1	‘	23	‘
GIY83410.1	*Caerostris darwini*	1	‘
GIY83411.1	′	2	‘
GIY83412.1	′	3	‘
GIY83413.1	′	4	‘
GIY50143.1	‘	5	‘
GIX85174.1	‘	6	‘
GIY55999.1	‘	7	Found via BLAST
GIX69490.1	*Caerostris extrusa*	1	Labeled AChE in NCBI database
GIX69495.1	‘	2	‘
GIX69498.1	‘	3	‘
GIY20531.1	‘	4	‘
GIY14843.1	‘	5	‘
GIY14844.1	‘	6	‘
GIX73517.1	‘	7	‘
GIX73520.1	‘	8	‘
GIX73522.1	‘	9	‘
KAF8767854.1	*Argiope bruennichi*	1	Found via BLAST
XP_035208789.1	*Stegodyphus dumicola*	1	‘
XP_035206450.1	‘	2	‘
KFM58795.1	*Stegodyphus mimosarum*	1	‘
KFM57104.1	‘	2	‘
UXX52860.1	*Pardosa astrigera*	1	‘
KAG8184384.1	*Oedothorax gibbosus*	1	‘
WBW70144.1	*Lampona murina*	1	‘

List of spider AChE amino acid sequences analyzed in this study. The third column shows the number assigned to that sequence in this study for the purpose of labeling the sequence on this study’s phylogenetic trees.

Table 2. Non-Spider AChE Amino Acid Sequences

NCBI Accession	Species	Assigned Number	Description
CAA27169.1	*Tetronarce californica*	1	Vertebrate AChE
CAA53080.1	*Caenorhabditis elegans*	1	Nematode ace-1
AAC14016.2	‘	2	Nematode ace-2
AAC14022.3	‘	3	Nematode ace-3
AAC14017.1	‘	4	Nematode ace-4
ABB89946.1	*Blattella germanica*	1	Insect ace1
ABB89947.1	‘	2	Insect ace2
KAG1667819.1	*Nymphon striatum*	1	Sea spider AChE-like
KAG1667818.1	‘	2	‘
KAG1656493.1	‘	3	‘
KAG1653309.1	‘	4	‘
KAG1653306.1	‘	5	‘
KAG1653307.1	‘	6	‘
KAG1653305.1	‘	7	‘
KAG1653308.1	‘	8	‘
KAG1656492.1	‘	9	‘
KAG1696623.1	‘	10	‘
KAG1656491.1	‘	11	‘
KAG1656490.1	‘	12	‘
KAG1682984.1	‘	13	‘
KAG1655484.1	‘	14	‘
KAG1658379.1	‘	15	‘
KAG1660197.1	‘	16	‘
KAG1650495.1	‘	17	‘
KAG1700081.1	‘	18	‘
KAG1700082.1	‘	19	‘
KAG1650494.1	‘	20	‘
KAG1650493.1	‘	21	‘
KAG1655485.1	‘	22	‘
KAG1655442.1	‘	23	‘
KAG1650497.1	‘	24	‘
KAG1650502.1	‘	25	‘
KAG1650500.1	‘	26	‘
KAG1650496.1	‘	27	‘
KAG1650498.1	‘	28	‘
KAG1650501.1	‘	29	‘
KAG1655483.1	‘	31	‘
KAG1655482.1	‘	32	‘
KAG1650499.1	‘	33	‘
KAG1653310.1	‘	34	‘

NCBI Accession	Species	Assigned Number	Description
KAG1692615.1	‘	35	‘
KAG1692614.1	‘	36	‘
KAG1692613.1	‘	37	‘
KAG1655509.1	‘	38	‘
KAG1668974.1	‘	39	‘
KAG1684637.1	‘	40	‘
KAG1668972.1	‘	41	‘
KAG1684635.1	‘	42	‘
KAG1668976.1	‘	43	‘
XP_022237259.1	*Limulus polyphemus*	1	Horseshoe crab AChE-like
XP_013793998.1	‘	2	‘
XP_022253295.1	‘	3	‘
XP_013780383.2	‘	4	‘
XP_013780687.2	‘	5	‘
XP_022258371.1	‘	6	‘
XP_022257289.1	‘	7	‘
XP_013790879.1	‘	8	‘
XP_022258739.1	‘	9	‘
XP_022258738.1	‘	10	‘
XP_022257792.1	‘	11	‘
XP_013779867.1	‘	12	‘
XP_013789679.1	‘	13	‘
XP_022256811.1	‘	14	‘
XP_013790869.1	‘	15	‘
XP_013793463.1	‘	16	‘
XP_013772240.1	‘	17	‘
XP_013783727.2	‘	18	‘
XP_022255593.1	‘	19	‘
XP_022255587.1	‘	20	‘
XP_013783581.1	‘	21	‘
XP_013774570.2	‘	22	‘
XP_013794815.1	‘	23	‘
XP_013791374.2	‘	24	‘
XP_022250755.1	‘	25	‘
XP_022250742.1	‘	26	‘
XP_022250749.1	‘	27	‘
XP_022248275.1	‘	28	‘
XP_013779919.2	‘	29	‘
XP_022242428.1	‘	30	‘
XP_013782247.2	‘	31	‘

NCBI Accession	Species	Assigned Number	Description
XP_022250761.1	‘	32	‘
XP_022241701.1	‘	33	‘
XP_022242429.1	‘	34	‘
XP_022248271.1	‘	35	‘
XP_013779866.1	‘	36	‘
XP_022250885.1	‘	37	‘
XP_022241668.1	‘	38	‘
XP_022241664.1	‘	39	‘
XP_022240830.1	‘	40	‘
XP_022240829.1	‘	41	‘
XP_013779898.1	‘	42	‘
XP_013772241.1	‘	43	‘
XP_022255609.1	‘	44	‘
XP_022254096.1	‘	45	‘
XP_023240877.1	*Centruroides sculpturatus*	1	Scorpion AChE-like
XP_023233519.1	‘	2	‘
XP_023240592.1	‘	3	‘
XP_023220570.1	‘	4	‘
XP_023220569.1	‘	5	‘
XP_023220568.1	‘	6	‘
XP_023241636.1	‘	7	‘
XP_023220572.1	‘	8	‘
XP_023235245.1	‘	9	‘
XP_023235259.1	‘	10	‘
XP_023243542.1	‘	11	‘
XP_023235261.1	‘	12	‘
XP_023235260.1	‘	13	‘
XP_023241635.1	‘	14	‘
XP_023220578.1	‘	15	‘
XP_023235243.1	‘	16	‘
XP_023242325.1	‘	17	‘
XP_023235246.1	‘	18	‘
XP_023236334.1	‘	19	‘
XP_023236342.1	‘	20	‘
XP_023241644.1	‘	21	‘
XP_023240465.1	‘	22	‘
XP_023236332.1	‘	23	‘
XP_023236335.1	‘	24	‘
XP_023236336.1	‘	25	‘
XP_023236323.1	‘	26	‘
XP_023236339.1	‘	27	‘

NCBI Accession	Species	Assigned Number	Description
XP_023236322.1	‘	28	‘
XP_023236319.1	‘	29	‘
XP_023235238.1	‘	30	‘
XP_023236338.1	‘	31	‘
XP_023210705.1	‘	32	‘
XP_023210704.1	‘	33	‘
XP_023232943.1	‘	34	‘
XP_023242075.1	‘	35	‘
XP_023217965.1	‘	36	‘
XP_023236325.1	‘	37	‘
XP_023210702.1	‘	38	‘
XP_023242067.1	‘	39	‘
XP_023221289.1	‘	40	‘
XP_023222471.1	‘	41	‘
XP_023222476.1	‘	42	‘
XP_023238926.1	‘	43	‘
XP_023238906.1	‘	44	‘
XP_023230746.1	‘	45	‘
XP_023236328.1	‘	46	‘
XP_023210237.1	‘	47	‘
XP_023220492.1	‘	48	‘
XP_023216731.1	‘	49	‘
XP_023216729.1	‘	50	‘
ADK12702.1	*Tetranychus urticae*	1	Mite AChE
USH09479.1	*Rhipicephalus microplus*	1	Tick AChE1
ASS83170.1	‘	2	Tick AChE2
AOA32870.1	‘	3	Tick AChE3

List of non-spider AChE amino acid sequences analyzed in this study. The third column shows the number assigned to that sequence in this study for the purpose of labeling the sequence on this study’s phylogenetic trees.

Table 3. Non-AChE Amino Acid Sequences

NCBI Accession	Species	Assigned Number	Description
AAP21002.1	*Drosophila melanogaster*	Est-6	Insect esterase

One non-AChE amino acid sequence was included in an analysis in this study. The third column shows the number assigned to that sequence in this study for the purpose of labeling the sequence on this study's phylogenetic trees.

References

Agozzino, L., & Dill, K. A. (2018). Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proceedings of the National Academy of Sciences*, *115*(37), 9092–9097. https://doi.org/10.1073/pnas.1810194115

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/s0022-2836(05)80360-2

Ballesteros, J. A., Santibáñez-López, C. E., Baker, C. M., Benavides, L. R., Cunha, T. J., Gainett, G., Ontano, A. Z., Setton, E. V., Arango, C. P., Gavish-Regev, E., Harvey, M. S., Wheeler, W. C., Hormiga, G., Giribet, G., & Sharma, P. P. (2022). Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msac021

Ballesteros, J. A., & Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Systematic Biology*, *68*(6), 896–917. https://doi.org/10.1093/sysbio/syz011

Battelle, B.-A., Ryan, J. F., Kempler, K. E., Saraf, S. R., Marten, C. E., Warren, W. C., Minx, P. J., Montague, M. J., Green, P. J., Schmidt, S. A., Fulton, L., Patel, N. H., Protas, M. E., Wilson, R. K., & Porter, M. L. (2016). Opsin repertoire and expression patterns in horseshoe crabs: Evidence from the genome of *Limulus polyphemus* (Arthropoda: Chelicerata). *Genome Biology and Evolution*, *8*(5), 1571–1589. https://doi.org/10.1093/gbe/evw100

Boc, A., Diallo, A. B., & Makarenkov, V. (2012). T-rex: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Research*, *40*(W1), W573–W579. https://doi.org/10.1093/nar/gks485

Carlson, D. E., & Hedin, M. (2017). Comparative transcriptomics of Entelegyne spiders (Araneae, Entelegynae), with emphasis on molecular evolution of orphan genes. *PloS One*, *12*(4), e0174102. https://doi.org/10.1371/journal.pone.0174102

Dohrmann, M., & Wörheide, G. (2017). Dating early animal evolution using phylogenomic data. *Scientific Reports*, *7*(1), 3599. https://doi.org/10.1038/s41598-017-03791-w

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., & Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences, 102*(40), 14338–14343. https://doi.org/10.1073/pnas.0504070102

Engdahl, C. (2017). *Selective inhibition of acetylcholinesterase 1 from disease-transmitting mosquitoes: Design and development of new insecticides for vector control.* https://www.diva-portal.org/smash/get/diva2:1094489/FULLTEXT01.pdf

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. https://doi.org/10.1016/j.cub.2018.03.064

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic Zoology, 19*(2), 99–113.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Jeong, J.-H., Kim, H., Ryu, S., & Kim, W. (2020). The first pycnogonid draft genome of *Nymphon striatum*. *Frontiers in Ecology and Evolution*, *8*. https://doi.org/10.3389/fevo.2020.554164

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods, 14*, 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Kim, Y. H. & Lee, S. H. (2013). Which acetylcholinesterase functions as the main catalytic enzyme in the Class Insecta? *Insect Biochemistry and Molecular Biology, 43*(1), 47–53. https://doi.org/10.1016/j.ibmb.2012.11.004.

Kono, N., Nakamura, H., Mori, M., Yoshida, Y., Ohtoshi, R., Malay, A. D., Pedrazzoli Moran, D. A., Tomita, M., Numata, K., & Arakawa, K. (2021). Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk. *Proceedings of the National Academy of Sciences of the United States of America, 118*(31), e2107065118. https://doi.org/10.1073/pnas.2107065118

Kono, N., Ohtoshi, R., Malay, A. D., Mori, M., Masunaga, H., Yoshida, Y., Nakamura, H., Numata, K., & Arakawa, K. (2021). Darwin's bark spider shares a spidroin repertoire with *Caerostris extrusa* but achieves extraordinary silk toughness through gene expression. *Open Biology*, *11*(12), 210242. https://doi.org/10.1098/rsob.210242

Kulkarni, S., Kallal, R. J., Wood, H., Dimitrov, D., Giribet, G., & Hormiga, G. (2020). Interrogating genomic-scale data to resolve recalcitrant nodes in the spider tree of life. *Molecular Biology and Evolution*, *38*(3), 891–903. https://doi.org/10.1093/molbev/msaa251

Kwon, D.H., Clark, J.M., & Lee, S.H. (2010). Extensive gene duplication of acetylcholinesterase associated with organophosphate resistance in the two-spotted spider mite. *Insect Molecular Biology*, *19*, 195–204. https://doi.org/10.1111/j.1365-2583.2009.00958.x

Kwon, D. H., Im, J. S., Ahn, J. J., Lee, J.-H., Marshall Clark, J., & Lee, S. H. (2010). Acetylcholinesterase point mutations putatively associated with monocrotophos resistance in the two-spotted spider mite. *Pesticide Biochemistry and Physiology*, *96*(1), 36–42. https://doi.org/10.1016/j.pestbp.2009.08.013.

Labbé, P., Berthomieu, A., Berticat, C., Alout, H., Raymond, M., Lenormand, T., & Weill, M. (2007). Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. *Molecular Biology and Evolution*, *24*(4), 1056–1067. https://doi.org/10.1093/molbev/msm025

Leone, P., Comoletti, D., Taylor, P., Bourne, Y., & Marchot, P. (2010). Structure–function relationships of the α/β-hydrolase fold domain of neuroligin: A comparison with acetylcholinesterase. *Chemico-Biological Interactions*, *187*(1-3), 49–55. https://doi.org/10.1016/j.cbi.2010.01.030

Lin, X., Zhang, Y., Yang, B., Zhang, L., Chen, Y., & Liu, Z. (2022). Multiple acetylcholinesterases in *Pardosa pseudoannulata* brain worked collaboratively to provide protection from organophosphorus insecticides. *Ecotoxicology and Environmental Safety*, *248*, 114301. https://doi.org/10.1016/j.ecoenv.2022.114301

Lockridge, O., Quinn, D. M., & Radić, Z. (2018). Esterases. *Comprehensive Toxicology*, 277–307. https://doi.org/10.1016/b978-0-12-801238-3.01970-x

Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D., & Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nature Communications*, *10*(1), 2295. https://doi.org/10.1038/s41467-019-10244-7

Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools

services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279. https://doi.org/10.1093/nar/gkac240

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, *92*(1), 155–161. https://doi.org/10.1007/s12041-013-0212-8

Meng, X., Li, C., Xiu, C., Zhang, J., Li, J., Huang, L., Zhang, Y., & Liu, Z. (2016). Identification and biochemical properties of two new acetylcholinesterases in the pond wolf spider (*Pardosa pseudoannulata*). *PLoS One*, *11*(6), e0158011. https://doi.org/10.1371/journal.pone.0158011

Meng, X., Xu, X., Bao, H., Wang, J., & Liu, Z. (2017). Characterization of the fifth putative acetylcholinesterase in the wolf spider, *Pardosa pseudoannulata*. *Molecules*, *22*(7), 1118. https://doi.org/10.3390/molecules22071118

Meng, X., Zhang, Y., Bao, H., & Liu, Z. (2015). Sequence analysis of insecticide action and detoxification-related genes in the insect pest natural enemy *Pardosa pseudoannulata*. *PLoS One*, *10*(4), e0125242. https://doi.org/10.1371/journal.pone.0125242

Meyer, W., & Idel, K. (1977). The distribution of acetylcholinesterase in the central nervous system of jumping spiders and wolf spiders (Arachnida, Araneida: Salticidae et Lycosidae). *The Journal of Comparative Neurology*, *173*(4), 717–743. https://doi.org/10.1002/cne.901730406

Meyer, W., & Pospiech, B. (1977). The distribution of acetylcholinesterase in the central nervous system of web-building spiders (Arachnida, Araneae). *Histochemistry*, *51*(2-3), 201–208. https://doi.org/10.1007/bf00567224

Michálek, O., Walker, A.A., Šedo, O., Zdráhal, Z., King, G. F., & Pekár, S. (2022). Composition and toxicity of venom produced by araneophagous white-tailed spiders (Lamponidae: *Lampona sp*.). *Scientific Reports*, *12*, 21597. https://doi.org/10.1038/s41598-022-24694-5

Michalko, R., Pekár, S., Dul'a, M., Entling, MH. (2019). Global patterns in the biocontrol efficacy of spiders: A meta-analysis. *Global Ecology and Biogeography*, *28*: 1366–1378. https://doi.org/10.1111/geb.12927

Mladenović, M., Arsić, B., Stanković, N., Mihović, N., Ragno, R., Regan, A., Milićević, J., Trtić-Petrović, T., & Micić, R. (2018). The targeted pesticides as acetylcholinesterase inhibitors: Comprehensive cross-organism molecular modelling studies performed to anticipate the pharmacology of harmfulness to humans in vitro. *Molecules*, *23*(9), 2192. https://doi.org/10.3390/molecules23092192

Montella, I. R., Schama, R., & Valle, D. (2012). The classification of Esterases: An important gene family involved in insecticide resistance - a review. *Memórias Do*

Instituto Oswaldo Cruz, *107*(4), 437–449. https://doi.org/10.1590/s0074-02762012000400001

National Center for Biotechnology Information. (n.d.). U.S. National Library of Medicine. Retrieved November 11, 2022, from https://www.ncbi.nlm.nih.gov/

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274. https://doi.org/10.1093/molbev/msu300

Ontano, A. Z., Gainett, G., Aharon, S., Ballesteros, J. A., Benavides, L. R., Corbett, K. F., Gavish-Regev, E., Harvey, M. S., Monsma, S., Santibáñez-López, C. E., Setton, E. V., Zehms, J. T., Zeh, J. A., Zeh, D. W., & Sharma, P. P. (2021). Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Molecular Biology and Evolution*, *38*(6), 2446–2467. https://doi.org/10.1093/molbev/msab038

Pekár, S. (2012). Spiders (Araneae) in the pesticide world: An ecotoxicological review. *Pest Management Science*, *68*(11), 1438–1446. https://doi.org/10.1002/ps.3397

Pezzementi, L., Nachon, F., & Chatonnet, A. (2011). Evolution of acetylcholinesterase and butyrylcholinesterase in the vertebrates: An atypical butyrylcholinesterase from the medaka *Oryzias latipes*. *PLoS One*, *6*(2), e17396. https://doi.org/10.1371/journal.pone.0017396

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., Buffry, A. D., Chao, H., Dinh, H., Doddapaneni, H. V., Dugan, S., Eibner, C., Extavour, C. G., Funch, P., Garb, J., … McGregor, A. P. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology*, *15*(1). https://doi.org/10.1186/s12915-017-0399-x

Sharma, P. P., Schwager, E. E., Extavour, C. G., & Wheeler, W. C. (2014). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proceedings of the Royal Society B*, *281*(1792), 20140661. https://doi.org/10.1098/rspb.2014.0661

Soreq, H., & Seidman, S. (2001). Acetylcholinesterase — new roles for an old actor. *Nature Reviews Neuroscience*, *2*(4), 294–302. https://doi.org/10.1038/35067589

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Turk, E., Čandek, K., Kralj-Fišer, S., & Kuntner, M. (2020). Biogeographical history of golden orbweavers: Chronology of a global conquest. *Journal of Biogeography*, 47. 10.1111/jbi.13838.