Hemocyanin gene family evolution in spiders (Araneae), with implications for phylogenetic relationships and divergence times in the infraorder Mygalomorphae

James Starrett a,⁎, Marshal Hedin b, Nadia Ayoub c, Cheryl Y. Hayashi a

a Department of Biology, University of California, Riverside, CA 92521, USA
b Department of Biology, San Diego State University, San Diego, CA 92182, USA
c Department of Biology, Washington and Lee University, Lexington, VA 24450, USA

A R T I C L E   I N F O

Article history:
Accepted 15 April 2013
Available online 28 April 2013

Keywords:
Gene duplication
Gene tree parsimony
Gene tree reconciliation
Molecular clock
Respiration
Supermatrix

A B S T R A C T

Hemocyanins are multimeric copper-containing hemolymph proteins involved in oxygen binding and transport in all major arthropod lineages. Most arachnids have seven primary subunits (encoded by paralogous genes α–g), which combine to form a 24-mer (4 × 6) quaternary structure. Within some spider lineages, however, hemocyanin evolution has been a dynamic process with extensive paralog duplication and loss. We have obtained hemocyanin gene sequences from numerous representatives of the spider infraorders Mygalomorphae and Araneomorphae in order to infer the evolution of the hemocyanin gene family and estimate spider relationships using these conserved loci. Our hemocyanin gene tree is largely consistent with the previous hypotheses of paralog relationships based on immunological studies, but reveals some discrepancies in which paralog types have been lost or duplicated in specific spider lineages. Analyses of concatenated hemocyanin sequences resolved deep nodes in the spider phylogeny and recovered a number of clades that are supported by other molecular studies, particularly for mygalomorph taxa. The concatenated data set is also used to estimate dates of higher-level spider divergences and suggests that the diversification of extant mygalomorphs preceded that of extant araneomorphs. Spiders are diverse in behavior and respiratory morphology, and our results are beneficial for comparative analyses of spider respiration. Lastly, the conserved hemocyanin sequences allow for the inference of spider relationships and ancient divergence dates.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Hemocyanins are copper-containing hemolymph proteins that bind oxygen and facilitate oxygen transport in all major arthropod lineages, including Chelicerata, Myriapoda, and Pancrustacea (Burmester, 2002, 2004; Ertas et al., 2009; Hagner-Holler et al., 2004; Kusche and Burmester, 2001; Rehm et al., 2012), as well as the arthropod relative, Onychophora (Kusche et al., 2002). Arthropod hemocyanins are part of a protein superfamily that is divided into two main molecular lineages (Burmester, 2001) – one lineage consists of arthropod phenoloxidases, involved in cuticle sclerotization, whereas the other lineage is composed of hemocyanins and proteins derived from hemocyanins (e.g., crustacean pseudohemocyanins, insect hexamerins, etc.; Fig. 1a). Hemocyanins are hypothesized to have evolved during the Cambrian, in association with an increase in arthropod body size and a coincident need for more efficient oxygen transport (Burmester, 2001; Kusche et al., 2002).

The primary structure of arthropod hemocyanins includes multiple oxygen-binding subunits. These subunits are approximately 620–660 amino acids in length and are encoded by different paralogous members of the hemocyanin gene family (Burmester, 2001). Depending on species, six related or identical subunits combine to form either hexameric (1 × 6) or multi-hexameric (2–8 × 6) quaternary structures. Beyond these broad generalizations, it is clear that subunit diversification and evolution of quaternary structure have followed different pathways within the major arthropod lineages (reviewed in Burmester, 2002). In arachnids, immunological studies have been used to identify seven primary subunits (paralogous genes α–g), which combine to form a 24-mer (4 × 6) quaternary structure. This arrangement likely represents the ancestral condition in arachnids (Markl, 1986; Markl et al., 1986). A combination of immunological, electron microscopic, and
DNA sequence studies have been used to study hemocyanin evolution in the arachnid order Araneae (spiders), and again, the seven subunit and 24-mer configuration appears ancestral (Rehm et al., 2012). For example, complementary DNA (cDNA) studies have shown that the distant relatives, Euryelpma (a tarantula) and Nephila (an orb-weaver), share six of the seven related hemocyanin subunits (Fig. 1b). The subunits are genetically distant, and the estimated divergences among subunits (550–420 MYA) predate the divergence of spiders themselves (Averdam et al., 2003; Rehm et al., 2012; Voit et al., 2000; Fig. 1b). Immunological and electron microscopic studies conducted on a broader sample of spider families (40 species from 25 families) confirm that a 4 × 6-mer configuration is most common (Markl, 1986; Markl et al., 1986).

Despite this apparent conservation in paralog representation and quaternary structure, other evidence indicates that hemocyanin evolution has sometimes been a dynamic process within Araneae. For example, the genus Cupiennius (a wandering spider), a distant relative of Nephila, shows a dramatic shift in protein evolution (Ballweber et al., 2002). All but one subunit appears to have been lost, but this subunit (g) has subsequently undergone rounds of duplication to result in at least six copies (Fig. 1b). Additional evidence suggests that similarly dramatic changes in subunit composition and quaternary structure have independently occurred in other spider lineages (Kuwada and Sugita, 2000; Markl, 1986; Markl et al., 1986).

The combination of both stasis and change seen within spider hemocyanins invites two avenues of research. First, characterization of hemocyanin gene family structure across the spider phylogeny may help reveal the factors that have led to diverse lifestyles, body sizes, and respiratory systems in different spider groups. Second, we utilize hemocyanin sequences to infer the phylogenetic relationships among major groups of Araneae. Although spiders represent one of the most diverse orders of arthropods, our understanding of spider phylogeny at many hierarchical levels remains uncertain (Agnarsson et al., 2012; Blackledge et al., 2009; Bond et al., 2012; Coddington and Levi, 1991; Miller et al., 2010; Regier et al., 2010; Ubick et al., 2005). Additional molecular phylogenetic data can only help in this regard, and available evidence suggests that spider hemocyanins evolve slowly enough to retain deep phylogenetic signal (see Averdam et al., 2003; Ballweber et al., 2002; Rehm et al., 2012; Voit et al., 2000).

1.1. Mygalomorph spiders

Mygalomorph spiders, which include the tarantulas, trapdoor spiders, and other less well-known groups, represent one of three main spider lineages (Fig. 2). Although mygalomorphs retain some features that are plesiomorphic in spiders (e.g., two pairs of book lungs), several characters support mygalomorph monophyly, and this monophyly has not been seriously questioned (see Platnick and Gertsch, 1976; Raven, 1985). Current estimates of mygalomorph diversity place roughly 2700 species into 326 genera and 16 families (Platnick, 2012). Mygalomorphs are essentially worldwide in distribution, with centers of generic diversity in all tropical regions as well as temperate austral areas of South America, southern Africa, and Australasia (Platnick, 2012; Raven, 1985).

Inferring relationships among mygalomorph taxa has been a considerable challenge for systematists. Mygalomorphs are notorious for conservative, and often homoplasic, patterns of morphological evolution making it difficult to determine the placement of many families, as well as larger clades (Goloboff, 1993; Raven, 1985). In addition, the monophyly of many families is based on few morphological synapomorphies. Recent studies based on molecular data (e.g., elongation factor-1 gamma (EF-1γ); Ayoub et al., 2007; ribosomal DNA; Bond and Hedin, 2006; Hedin and Bond, 2006; combined ribosomal DNA

Fig. 1. (a) Evolutionary relationships of proteins in the arthropod hemocyanin superfamily, including estimated times of molecular divergence (adapted from Fig. 2 of Burmester, 2002). (b) Evolutionary relationships of hemocyanin subunits in spiders, including estimated times of molecular divergence (adapted from Fig. 7 of Averdam et al., 2003).

Fig. 2. Summary phylogeny of spiders (Araneae) based on the hypotheses of Agnarsson et al. (2012), Blackledge et al. (2009), Bond et al. (2012), Coddington et al. (2004), and Miller et al. (2010).
and EF-1 τ (Bond et al., 2012) have contributed to a better understanding of mygalomorph relationships and contradicted monophyly of multiple families, including Cyrtarachniidae, Dipluridae, Hexathelidae, and Ctenizidae (Fig. 2). However, pervasive conflict among these gene trees is apparent, highlighting the need for sampling additional phylogenetic markers to corroborate hypotheses of higher-level relationships.

In this paper, we sample mygalomorph and araneomorph taxa to explore patterns of hemocyanin molecular evolution and infer phylogenetic relationships in spiders. Our investigation supports hypotheses of a complex history of hemocyanin evolution with both extreme conservation and episodic lineage specific duplication and loss. Additionally, we apply supermatrix and gene tree parsimony methods to infer higher-level spider phylogenetic relationships using hemocyanin sequences. Finally, utilizing relaxed clock methods, we estimate higher-level spider divergence dates from hemocyanin sequences, and compare our estimates to previously hypothesized molecular clock dates.

2. Materials and methods

2.1. Taxon sampling

We collected new hemocyanin gene sequences from seventeen genera representing nine mygalomorph families (Fig. 2, Table 1). Collection information for most specimens is shown in Table S1. The families include both members of Atypoida (a monophyletic group within Atypoidea comprised of Antrodiaetidae and Atypidae; Raven, 1985), multiple genera representing the hypothesized paraphyletic family Dipluridae, and six families sampled broadly across the Bipectina clade (Hedin and Bond, 2006). Within Bipectina, we sampled multiple genera representing the likely non-monophyletic families, Cyrtarachniidae and Ctenizidae. In addition, we downloaded sequences from GenBank for the theraphosids “Euryelpma californium” (= Aphonopelma hentzi; Nentwig, 2012; Voit and Feldmaier-Fuchs, 1990; Voit et al., 2000; Voll and Voit, 1990) and Acanthoscurria gomesiana (Lorenzini et al., 2006).

To compare patterns of hemocyanin evolution across araneomorphs and mygalomorphs, we broadly sampled across the araneomorph phylogeny (Fig. 2, Table 1; Blackledge et al., 2009; Coddington and Levi, 1991; Coddington et al., 2004; Dimitrov et al., 2012; Wood et al., 2012). We sequenced hemocyanin gene fragments from araneomorphs representing ten genera from ten families. Specifically, we sampled the palcercibrillate representative Hypochilus, two members of Haplagnayna (Kukulcania and Diguetia), three Araneoida (a monophyletic group within Orbiculariae; Gasteracantha, Nephila, and Nesticus), and four members of the retrotralateral tibial apophysis (RTA) clade (Habronattus, Tenggella, Allocosa, and Cupiennius). Cupiennius and Nephila sequences were obtained from GenBank (Averdam et al., 2003; Ballweber et al., 2002). We rooted the spider hemocyanin trees with the inferred DNA coding sequence (back-translation using MacVector 7.2 Accelrys, Inc., San Diego, CA) for the homologous region of the Aa6 hemocyanin subunit from the scorpion Androctonus australis (Buzy et al., 1995).

The assignment of the Androctonus sequence to a specific ortholog group is currently ambiguous (Averdam et al., 2003; Rehm et al., 2012); however, the sequence does not fall within spiders for any ortholog group, and thus serves as a suitable outgroup. An additional analysis (not shown) using Endo spirona (Chelicera, Pantopoda) hemocyanin sequence EspHCl (GenBank accession FR865911) as the orthogroup resulted in the same assignment of spider hemocyanin sequences to ortholog groups as when Androctonus Aa6 was used as the outgroup.

2.2. Hemocyanin sequences

For the amplification of hemocyanin genes from the taxa listed in Table 1, genomic DNAs were extracted using the DNeasy tissue kit (Qiagen, Valencia, CA). We amplified these DNAs with primers designed partly on the gene structure of “Euryelpma” subunit e, which is divided into nine short exons (129–476 bp) separated by longer (2.1–143 kbp) introns (Voll and Voit, 1990). Because of the unpredictability of intron sizes, we targeted our primers to amplify within exon 4. This exon is one of the largest exons (376 bp) and includes coding sequence for one of the two highly-conserved copper-binding sites of the hemocyanin complex (Voit et al., 2000; Voll and Voit, 1990). Sets of forward and reverse primers were designed from published Cupiennius (Ballweber et al., 2002) and “Euryelpma” sequences (Voit et al., 2000).

Forward and reverse primer cocktails were made by combining equal molar amounts of the individual primers (Table S2). PCR amplification was done in a 50 μl reaction volume with 4 μM of each primer cocktail, 0.5 mM each dNTP, 67 mM Tris, 3 mM MgCl2, 16.6 mM (NH4)2SO4, and 1 unit of recombinant Taq DNA polymerase (Invitrogen, Carlsbad, CA). Cycling conditions consisted of 94 °C for 1 min, then 50 cycles of 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 1 min, and a final 72 °C for 5 min.

PCR products were cloned into pCR2.1-TOPO plasmids (Invitrogen, Carlsbad, CA) and electroporated into TOP10 Escherichia coli cells (Invitrogen). For each species, recombinant colonies were screened in sets of 24 and ~15 colonies with inserts of the target size were sequenced. Sequencing was done with the T7 universal primer at the U.C. Riverside Genomics Core Instrumentation Facility. DNA sequencing results were edited and translated with MacVector 7.2.

Hemocyanin sequences obtained from two taxa, Eurgus chiseous and Aphonopelma seemanni, were found in cdNA libraries that were constructed for the characterization of silk fibroin transcripts. Methods used to generate the cdNA library for Eurgus were described in Gately et al. (2001). cdNA library construction methods for A. seemanni followed Garb et al. (2007). Hemocyanin homologs from EST sequences were identified using BlastX against the NCBI nr protein database (Altschul et al., 1997).

DNA and amino acid sequences were aligned with Clustal W (Thompson et al., 1994) implemented in MacVector and manually adjusted. The length of the DNA alignment was 358 bp. All of the sequences were of the same length except for a three base insertion in putative orthologs of hemocyanin subunit f.

2.3. Phylogenetic analyses

Parsimony and Bayesian phylogenetic analyses were conducted on multiple hemocyanin data sets (described in Sections 2.3.1 and 2.3.2). Parsimony analyses were conducted using heuristic searches with 1000 random taxon addition replicates and tree-bisection–reconnection (TBR) branch swapping, as implemented in PAUP* 4.0b8-b10 (Swofford, 2002). Clade support was evaluated using nonparametric bootstrapping (Felsenstein, 1985), based on analyses comprising 1000 pseudoreplicates (heuristic TBR branch-swapping, ten random taxon addition replicates per pseudoreplicate).

A combination of unpartitioned and partitioned Bayesian analyses (Brandley et al., 2005; Huelsenbeck and Ronquist, 2001; Huelsenbeck et al., 2002; Miller et al., 2002) was conducted on hemocyanin matrices using MrBayes v. 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Model choices were determined by MrModeltest v. 2.1 (Nylander, 2004; all models selected are shown in Table S3) using the Akaike Information Criterion (AIC) following Posada and Buckley (2004). Two searches were run simultaneously for at least 5 million generations, with trees and parameters sampled from four MCMC chains every 100th generation. Partitions (described in Sections 2.3.1 and 2.3.2) were unlinked and the substitution rates of evolution among partitions were allowed to vary. Analyses were considered complete when the standard deviation of split frequencies fell below 0.01 (Ronquist et al., 2005). The first forty percent of samples were treated as burnin and discarded. Bayesian factor tests comparing harmonic means were used to determine which partitioning scheme was justified (Brandley et al., 2005; Brown and Lemmon, 2007; Kass and Raftery, 1995).
2.3.1. Identification of hemocyanin paralogs

Initially, phylogenetic analyses were conducted on a matrix consisting of all hemocyanin sequences generated. Analyses were then conducted on a reduced hemocyanin data set that excluded allelic variants and putative PCR errors (see Results, Section 3.1) to determine the gene tree structure. Analyses were then conducted on split-humpless Bayesian analyses where the scorpion sequence was used as an outgroup (Table 2). In the Bayesian analyses, the concatenated data set was analyzed as unpartitioned, partitioned by codon position (3 partitions), partitioned by paralog (7 partitions), and partitioned by codon position and paralog (21 partitions).

2.3.2. Inference of species relationships

We also used hemocyanin sequences to infer higher-level phylogenetic relationships. First, analyses were conducted on the seven ortholog groups (referred to as individual analyses), each using the scorpion sequence as an outgroup (Table 2). In the Bayesian analyses, the concatenated data set was analyzed as unpartitioned, partitioned by codon position (3 partitions), partitioned by paralog (7 partitions), and partitioned by codon position and paralog (21 partitions).

2.3.3. Gene tree parsimony

We used the program Duptree (Wehe et al., 2008) to conduct gene tree parsimony bootstrap analyses (Cotton and Page, 2002; Page, 2000; Page and Charleston, 1997;Slowinski and Page, 1999). Duptree takes binary gene trees and uses Subtree Pruning and Regrafting with a fast heuristic algorithm to search for the species tree that requires the fewest duplication events (Wehe et al., 2008). Duptree does not count gene loss events in the reconciliation score. However, PCR is not an exhaustive method for determining gene presence or absence, and thus considering duplication events only was appropriate for our data set.

Searches were conducted with two different sets of input trees. For the first analysis, a codon partitioned Bayesian analysis was conducted on the global data set that included a single representative of the RTA clade (Cupiennius g). One hundred random post-burnin trees from this analysis were used as the binary input trees for Duptree. For the second analysis, one hundred random post-burnin trees from each partitioned by codon position Bayesian analysis of the individual ortholog groups were used as the binary input trees. Again, a single representative of the RTA-clade (Cupiennius g) was used. Duptree analyses were conducted on both data sets to determine the sensitivity of gene tree parsimony to differences in orthologous relationships inferred from the global data set versus the individual ortholog groups. In both analyses, the Duptree search was conducted with 1000 starting trees from a leaf adding heuristic, which were queued to start a full hill climbing heuristic. The maximum number of optimal species trees was retained (100,000 trees from each partitioned by codon position). In the Bayesian analyses, the concatenated data set was analyzed as unpartitioned, partitioned by codon position (3 partitions), partitioned by paralog (7 partitions), and partitioned by codon position and paralog (21 partitions).

2.4. Divergence dates

The hemocyanin gene family is highly conserved and evolves in a nearly clocklike manner within Chelicerata (Burmester, 2001, 2002), suggesting that it may be useful for dating ancient nodes within spiders (Rehm et al., 2012). We obtained a fully bifurcating input tree for use in molecular clock analyses by conducting a Bayesian analysis.
(partitioned by codon position) on the concatenated data set excluding taxa represented by a single paralog (i.e., Aliatypus, Stasimopus, Diguetia, Nesticus, all RTA clade representatives). In addition, the scorpion sequence was not included because outgroups are not necessary for dating with the method we utilized (Inoue et al., 2010; below). Instead, the resulting fully bifurcating consensus tree (see Results, Section 3.5) was rooted with the remaining araneomorph spiders. This tree did not conflict with phylogenies that included all taxa.

The Baseml program in PAML v. 4.4c (Yang, 1997, 2007) was used to determine the likelihood values of a global molecular clock model and an independent rates model for the concatenated data set. A likelihood ratio test comparing the two models rejected a global molecular clock (df = 20, p < 0.001). Therefore, Bayesian analyses using MCMCtree in the PAML package were conducted. Sequence data from the three codon positions were treated as separate partitions. The analysis was run with a model that assumes correlated rates among branches, as well as with a model that allows independent rates among branches (Rannala and Yang, 2007; Yang, 2006; Yang and Rannala, 2006). We estimated the parameters (shape and scale) of the gamma distribution of the overall substitution rate prior (μ) using Baseml estimates of the substitution rate for each partition, based on five fixed calibration points (shape = 0.69, scale = 5.38). Two gamma distributed priors were tested for σ², which specifies how variable the substitution rate is among branches (shape = 1, scale = 1 versus shape = 1, scale = 10; with the former specifying greater rate variation among branches). The HKY sequence model was used and the analysis was run with birth rate, death rate, and species sampling priors of 2, 2, and 0.1, respectively. Gamma priors for κ (the transition/transversion ratio) and α (shape parameter for among site rate variation) were left as default (Yang, 2006). Calibrations (see below) were treated as soft boundaries (i.e., 2.5% chance date falls beyond boundary; Inoue et al., 2010; Yang and Rannala, 2006). The first 20,000 iterations were discarded as
burnin, followed by 50,000 iterations sampled every five iterations. Analyses were run twice to ensure MCMC convergence, with negligible differences in posterior date estimates between the two runs of the same analysis.

Five fossil calibrations were used in our analysis, three of which were used in a molecular clock analysis based on the gene encoding elongation factor-1 gamma (EF-1γ; Ayoub and Hayashi, 2009; Ayoub et al., 2007). *Rosamsyge*, the oldest mygalomorph fossil from the mid Triassic, was used as a minimum age of 240 million years ago (MYA) for the earliest mygalomorph divergence (*Selden and Gall*, 1992). *Cretamygale*, a bipectin mygalomorph from the early Barremian (Cretaceous), was used as a minimum age of 130 MYA for the deepest bipectin node (*Selden*, 2002). Two fossil Atypoida spiders, *Cretacattyma* (*Antrodiaetidae*) and *Ambiortiphagus* (*Atypidae*), from the upper half of the Lower Cretaceous, were used to place a minimum age of 96 MYA for the antrodiaetid–atypid split (*Eskov and Zonstein*, 1990). The sole araneomorph fossil representative used in the analysis was the nephilid, *Nephele jurassica*, from the Middle Jurassic, which placed a minimum age of 165 MYA for the nephilid–araneid split (*Selden et al.*, 2011). A maximum age of 392 MYA for the root node in the tree was provided by *Attercorpus* from the Middle Devonian (*Selden et al.*, 1991). *Attercorpus* is a member of the Urraneidae, the hypothesized sister order to spiders, and is older than all known spider fossils (*Selden et al.*, 2008).

### 3. Results

#### 3.1. Inference of ortholog groups

Our initial global hemocyanin matrix included 110 unique clone sequences, plus 22 GenBank sequences (*n* = 132). Heuristic parsimony analysis of this data set (not shown) indicated that within the clearly distinguishable ortholog groups (see arguments below), multiple clones from the same species typically formed clades of closely related or near-identical sequences. We interpreted these sequences as representing either allelic variation or PCR error. For the subsequent phylogenetic analyses, we collapsed closely related sequences into a single consensus sequence using majority rule and ambiguity coding for ties, reducing the global matrix to 84 unique sequences. All data matrices and resulting phylogenetic trees (see below) are deposited in Dryad (doi:10.5061/dryad.105c0).

Bayes factor analysis revealed that partitioning by codon position fits the data significantly better than no partitioning (Table S4). Both partitioned (Fig. 3) and unpartitioned (not shown) Bayesian analyses of the global matrix yielded seven distinct clades with high posterior probabilities (PP ≥ 0.94; Fig. 3). These seven clades corresponded to the seven hemocyanin subunits (paralogous genes a–g) that have been well documented in *Euryypela* (*Voit et al.*, 2000), six of which have also been found in *Nephila* (*Averdam et al.*, 2003; see Fig. 1b). Each sequence clade included a subclade of mygalomorph sequences with a *Euryypela* paralog, and a suboclade of araneomorph sequences with a *Nephila* paralog (Fig. 3). The exception is subunit c, which apparently has been lost in *Nephila* (*Averdam et al.*, 2003), and was not found in any of the araneomorphs that we sampled. All hemocyanin sequences from spiders representing the RITA clade (*Habronattus*, *Tengella*, *Allocosa*, and *Cupiennius*) were restricted to the subunit g clade.

Relationships among the seven clades were unresolved or poorly supported with the exception of a strongly supported clade consisting of subunits b + c. The tree topology based on the parsimony analysis (Fig. S1) was largely consistent with that of the Bayesian analysis but less resolved. Relationships that conflicted under the two different optimality criteria occurred at nodes with low support.

#### 3.2. Species relationships inferred from individual ortholog groups

Based on the results of the global analyses, we generated seven matrices corresponding to the groups a–g (Table 2). For each data set, Bayesian analyses (Fig. 4) with codon partitioning had a significantly better fit to the data than no partitioning, based on Bayes factor analysis (Table S4). For most of the ortholog groups, the topology based on the partitioned Bayesian analysis was more resolved than that under the parsimony bootstrap (Fig. S2). Only in two ortholog groups did the Bayesian and parsimony bootstrap topologies conflict (Figs. 4, S2). However, the placement of the diplurid, *Ischnothele*, within *Bipectina* in the Bayesian analysis of paralog d versus its placement as sister to *Bipectina* in the parsimony analysis was not well supported (PP = 0.65, bootstrap percent (BP) = 51). Similarly, the Bayesian analysis of paralog f placed araneomorphs within Mygalomorphae with low support (PP = 0.76), in contrast to the parsimony analysis, which recovered a well-supported monophyletic Mygalomorphae (BP = 79).

Among the Bayesian analyses of the individual ortholog groups, conflict in species relationships occurred among a, d, e, and f. Conflict in relationships was mostly associated with nodes that have low support. One exception of this conflict between ortholog groups was in the placement of the diplurid, *Eurypelma* (Fig. 4). In the analysis of paralog f, *Eurypelma* falls within a strongly supported clade of bipectin spiders (PP = 0.99), whereas in paralog e, *Eurypelma* is sister to a strongly supported clade of *Bipectina* (PP = 1).

We also compared orthologous relationships recovered in Bayesian analyses of individual paralogs (Fig. 4) to those in the Bayesian analysis of the global data set (Fig. 3). Most orthologous relationships were identical across the individual and global data set analyses. Instances of conflict were associated with nodes with low support. One exception to this was that the Bayesian analysis of individual ortholog group g recovered a monophyletic Araneomorphae (PP = 0.99), whereas araneomorphs were not monophyletic in the corresponding g clade based on the global analysis (*Hypochilus* g was sister to a mygalomorph g clade; PP = 0.99).

#### 3.3. Species relationships inferred from concatenated paralogs

Despite discordance among individual gene trees, Bayesian analysis of the concatenated data set produced a tree resolved at twenty of
the twenty-six possible nodes, with strong support at twelve nodes (Fig. 5). Parsimony analysis of the concatenated data set was consistent with the Bayesian analyses but resulted in a largely unresolved topology (Fig. S3). For the Bayesian analyses, Bayes factor analysis indicated that partitioning the data by codon position was most appropriate (Table S4). However, topologies from different analyses under unpartitioned and different partitioning schemes were consistent, only differing in degree of resolution. All analyses recovered a monophyletic Mygalomorphae with high support (PP ≥ 0.96).

Within mygalomorphs, a monophyletic Atypoidina and a monophyletic Bipectina were both always recovered with high support (PP = 1). Diplurids were paraphyletic in each analysis with a clade of Euagrus plus Allothele being most closely related to Bipectina (PP ≥ 0.91). Within the Bipectina, theraphosids formed a clade that was sister to all others. Also, consistent across analyses was a clade that included “ancylotripines” (Homostola and Anclylotrypa) allied with the nemesiids, Acanthogonatus and Stamwellia.

A well-supported monophyletic Araneomorphae was recovered in Bayesian analyses for all partitioning schemes (PP ≥ 0.98). In each analysis, the paleocribellate, Hypochilus, was sister to all other araneomorphs. Relationships among the remaining araneomorphs were unresolved, with the exception of a weakly supported relationship between the nephilid, Nephila, and araneid, Gasteracantha.

3.4. Species relationships inferred from gene tree reconciliation

For the Duptree analyses of the gene trees based on the global and individual ortholog group data sets, the resulting bootstrap trees were largely unresolved (Figs. 6, S4). Both 50% majority-rule trees resulted in a similar number of resolved nodes (global = 11, individual = 10). Mygalomorph monophyly was recovered in the global analysis, but the placement of members of Atypoidina was unresolved in the Duptree analysis of the tree sets from the individual ortholog groups. Both Duptree analyses recovered a non-atypoid clade consisting of...
members of Bipectina plus Dipluridae, although with weak support (BP ≤ 66). Within this clade, most relationships were unresolved except for a weakly-supported clade consisting of Ancylotrypa plus nemesiids (global includes Homostola), a theraphosid clade with low support, and a diplurid clade also with low support (BP ≤ 67 for each of the three mentioned clades). It is notable that the bootstrap proportions were similar for these latter two clades given the high support (PP = 1) for theraphosid monophyly in the global and individual (a–g) ortholog analyses. By contrast, a diplurid clade was recovered only in the a ortholog group in the global (PP = 0.87) and individual (PP = 0.98) analyses. In the global analysis and ortholog group e individual analysis, diplurids were inferred to be polyphyletic. Furthermore, in the global analysis and ortholog group f individual analysis, the placement of Thelechoris was unresolved with respect to a clade consisting of the diplurids, Euagrus and Allothele, and a clade consisting of members of Bipectina.

Relationships among the araneomorph taxa were largely unresolved in the bootstrap trees arising from both Duptree analyses (Figs. 6, S4). In the global analysis, araneomorphs formed a polytomy and only a single relationship was recovered (Nephila plus Gasteracantha). Araneomorphs were also not monophyletic in the Duptree analysis of the individual ortholog groups, and relationships among taxa were unresolved except for a clade consisting of Nephila and Gasteracantha.

3.5. Molecular clock analysis

Bayesian estimates of mean node ages and 95% credibility intervals were similar between analyses with correlated rates among branches (Fig. 7, Table 3) versus independent rates among branches (not shown; estimates were within ~9 Myr difference in mean date estimates and within ~12 Myr for 95% CI range size). The posterior estimates of substitution rates (µ) for each of the codon positions were similar (within 0.01 substitutions/site) across analyses regardless of the priors used for σ², and whether rates among branches were correlated or independent (Table 3).

4. Discussion

4.1. Gene family evolution

The phylogenetic analysis of hemocyanin sequences from our broad taxonomic sample reveals that hemocyanin gene family evolution has been both static and dynamic in spiders. The seven clade structure recovered in our global analysis (Fig. 3) is consistent with the inferred ancestral arachnid condition of seven subunits that arose via six duplication events, as retained in the tarantula, Eurypelma (Burmester, 2001; Markl, 1986; Rehm et al., 2012; Voit et al., 2000). Lorenzini et al. (2006) identified a possible representative of an eighth paralog in the tarantula, Acanthoscurria gomesiana, most similar to the f paralog based on BLAST (Altschul et al., 1990); however, the partial transcript sequence did not overlap with our data set and was not included. Although we did not recover all seven paralogs in any single mygalomorph taxon, our data are consistent with relative stasis of the hemocyanin gene family in this group (Markl et al., 1986). For example, all paralogs are found in at least two mygalomorph species, although subunit b was restricted to theraphosids. Additionally, no paralog duplicates were detected in any mygalomorph taxa. This is inconsistent with the results of Kuwada and Sugita (2000), who, based on divergent N-terminal protein sequences, reported extensive paralog duplication and loss within various mygalomorph taxa. The limited coverage of paralogs for any particular mygalomorph species in our study could...
Fig. 7. Chronogram from Bayesian analysis using correlated rates. Tree topology is based on Bayesian analysis of the trimmed concatenated data set (species represented by only one paralog removed). Node numbers correspond to Table 3. Arrows (> and <) indicate minimum and maximum soft boundaries, respectively (see text for details). Dashed rectangles indicate 95% CI of the estimated date for each node.

In contrast to mygalomorph spiders, our results suggest that hemocyanin evolution has been dynamic in araneomorph spiders. Based on protein biochemical analysis and cDNA library data, Averdam et al. (2003) concluded that the c subunit is likely absent in Nephiia inaurata. We also did not detect c subunits in any araneomorph spider we sampled, which may be due to gene loss. However, absence of c paralogs may be due to incomplete sampling as our PCR method of data collection was not exhaustive. Improvements in primer design based on the growing number of published arachnid hemocyanins should allow for better recovery of hemocyanin paralogs with PCR in future studies. In Eurypelma, the c subunit acts as a linker molecule with the b subunit and is essential in forming the 4 × 6 quaternary structure (Markl et al., 1981, 1982; Van Bruggen et al., 1980). Despite this critical role for c in Eurypelma, many araneomorphs potentially achieve the same 4 × 6 quaternary structure without c. Given the high sequence conservation of hemocyanin paralogs, in some spider lineages it may be possible that different subunits can perform the same oxygen or subunit binding function as other subunits and compensate for paralog loss.

Spiders from the RTA clade exhibit the greatest deviation from the ancestral hemocyanin seven-subunit condition. Hemocyanin sequences from our sample of RTA clade members are consistent with those of Cupiennius, where all but subunit g appears to have been lost, and this subunit has subsequently duplicated. In the hemolymph of Cupiennius, hemocyanins may be found as single hexamers (1 × 6) or hexamer duplexes (2 × 6) rather than the larger 4 × 6 conformation observed in other spiders (Markl, 1980; Markl et al., 1976). The reduced quaternary structures may be a molecular synapomorphy for the RTA clade, as evidenced by the same crossed immunoelectrophoresis pattern shared among spiders representing twelve RTA clade families (Markl et al., 1986). The RTA clade of spiders is very speciose and ecologically diverse (Coddington and Levi, 1991). We have sampled species from RTA groups Lycosoidea, Dyschoroboidea, and Tengellidae, but much of the RTA clade remains unexplored for hemocyanin gene sequences (Agnarsson et al., 2012; Miller et al., 2010; Spagna and Gillespie, 2008). Future studies in which hemocyanin sequences are sampled from a greater phylogenetic breadth of RTA species that also represent ecological diversity will provide insight into the connection between hemocyanin duplication history, species diversity, and activity levels. In addition, Markl et al. (1986) showed that the haplogyne spider, Dysdera, exhibited a similar single hexameric hemocyanin composed of the same subunits as Cupiennius, and suggested that this is a complex molecular trait uniting haplogynes with the RTA clade. However, Rehm et al. (2012) propose that Dysdera lack hemocyanins entirely. The possession of the d subunit rather than the g subunit in our haplogyne sample, Diguetia, does not support a clade consisting of RTA plus haplogyne spiders. Further investigation is needed to determine the extent of hemocyanin loss in haplogyne spiders.

Evolution of the hemocyanin gene family may correlate with general patterns of morphological and behavioral evolution in spiders. Most mygalomorph species are bulky and live sedentary lifestyles, occupying a burrow for the duration of their lifetime (Foelix, 1996). Mygalomorphs also exhibit considerable conservatism in respiratory morphology; all mygalomorphs retain the ancestral condition of possessing two pairs of book lungs (Raven, 1985). The apparent stasis in hemocyanin evolution in mygalomorphs may reflect their conserved respiratory features and sedentary life histories.

In contrast to mygalomorphs, araneomorph spiders show considerable variation in behavior and respiratory morphology. Some araneomorph species live fossorial lifestyles similar to those of mygalomorphs. Others, such as orbicularians, expend energy to construct large aerial webs but then sit and wait until prey entangle in their webs (Blackledge et al., 2009; Foelix, 1996). In contrast, many species of the RTA clade live highly active, cursorial lifestyles (Coddington and Levi, 1991). Respiratory morphological evolution

---

**Table 3**

<table>
<thead>
<tr>
<th>Node</th>
<th>Soft bounds (MYA)</th>
<th>Posterior mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>~392</td>
<td>381.34 (329.95, 423.65)</td>
</tr>
<tr>
<td>2</td>
<td>394.30 (232.18, 365.15)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>255.71 (194.92, 322.95)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>&gt; 165</td>
<td>170.72 (136.75, 211.86)</td>
</tr>
<tr>
<td>5</td>
<td>&gt; 240</td>
<td>332.42 (270.81, 393.70)</td>
</tr>
<tr>
<td>6</td>
<td>237.66 (188.77, 297.90)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>209.85 (166.55, 264.08)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>&gt; 130</td>
<td>158.33 (126.56, 205.75)</td>
</tr>
<tr>
<td>9</td>
<td>146.90 (114.49, 191.86)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>130.15 (99.33, 171.37)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>123.91 (92.81, 164.64)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>109.83 (75.90, 151.90)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>67.26 (38.98, 104.10)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>21.25 (8.16, 39.50)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>111.95 (73.05, 156.98)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17.94 (10.31, 31.73)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>13.70 (7.49, 22.56)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>6.42 (1.93, 14.13)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>136.65 (87.92, 191.91)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>121.94 (68.31, 175.16)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>&gt; 96</td>
<td>117.76 (91.60, 172.38)</td>
</tr>
</tbody>
</table>

**Rates**

| μ1   | 0.0415 (0.0268, 0.0678) |
| μ2   | 0.0204 (0.0097, 0.0434) |
| μ3   | 0.2315 (0.1758, 0.3071) |
| σb1  | 0.1267 (0.0227, 0.4722) |
| σa2  | 0.5280 (0.0962, 1.4202) |
| σ13  | 0.0717 (0.0081, 0.2303) |
has been considerably more dynamic in araneomorphs as compared to mygalomorphs. With the exception of paleocribellates (e.g., *Hypochilus*), most araneomorph spiders have one pair of book lungs and the other pair has evolved into tubular tracheae that extend into the body (Foelix, 1996). The degree of tracheal branching varies from short tubes restricted to the opisthosoma to elaborately branched tracheae extending into the prosoma and extremities. The dramatic turnover of hemocyanins in araneomorphs, particularly in RTA clade spiders, may reflect the extensive diversity of behaviors and respiratory morphology in this group (Markl, 1986). Further investigation of the duplication history of hemocyanin paralogs and differences in oxygen binding and transport abilities should reveal whether there are selective advantages to the different hemocyanin structures.

### 4.2. Species tree analyses

Multi-gene families undoubtedly provide important information for inferring phylogenetic relationships, but there is ongoing debate on how best to utilize data from multiple loci to infer species relationships (Bininda-Emonds, 2004; Bull et al., 1993; de Queiroz and Gatesy, 2006; Edwards, 2008; Slowinski and Page, 1999). Few studies have compared results from supermatrix and gene tree parsimony methods applied to single data sets (Cotton and Page, 2003; Hartmann et al., 2012; Holton and Psani, 2010; McGowen et al., 2008; Simmons and Freudenstein, 2002). In our data set of hemocyanin sequences, there is considerable conflict in relationships among most of the ortholog groups. Gene loss and/or unsampled paralogs make our taxonomic sample sizes for each locus relatively small. In addition, selection of a proper outgroup becomes challenging in this situation where relationships among paralogs are unclear and a distant taxon must be selected for rooting purposes.

Individually, the single ortholog groups used here have limited ability to accurately infer species relationships (Figs. 4, S2). Hence, we generated species trees from the entire hemocyanin gene family using two different methods. Species trees resulting from analyses of the concatenated data (Figs. 5, S3) differ with species trees generated from gene tree parsimony analyses (Figs. 6, S4). Concatenation appears to be less sensitive to missing data and disagreement among paralogs, resulting in a more resolved tree than both gene tree parsimony analyses (i.e., from global hemocyanin trees and individual ortholog group trees). Both gene tree parsimony analyses are unable to recover a monophyletic Araneomorphae, and Mygalomorphae is only recovered in the analysis of the global trees, although with low support. Although trees based on concatenation are better resolved, we note that many relationships recovered are supported by data only from single paralogs and require further corroboration. Improved coverage of paralogs with more thorough taxonomical sampling will likely result in better-resolved trees with higher support values for both concatenation and gene tree parsimony methods. However, we note that our concatenated data set was limited in the sense that relationships among RTA clade spiders could not be assessed due to extensive duplication of the γ paralog, where ortholog groups could not be determined within this lineage.

Our results indicate that the hemocyanin gene family is phylogenetically informative, particularly for mygalomorphs. The concatenated analysis supports monophyletic Atypoidae and Bipectina clades, consistent with the results based on ribosomal DNA (rDNA) and morphology (Goloboff, 1993; Hedin and Bond, 2006). The family Dipluridae presents an interesting conflict between our concatenated and gene tree parsimony analyses. In concatenated analyses, diplurid representatives form a basal grade with respect to bipectin spiders as hypothesized by the studies of morphology and rDNA (Goloboff, 1993; Hedin and Bond, 2006). Phylogenetic analysis of concatenated hemocyanins placed the ischnotheline diplurids, *Ischnothelis* and *Thelechoris*, sister to a clade consisting of the euagrine diplurids, *Euaegrus* and *Allothela*, plus Bipectina, with moderate support. This relationship contrasts with those in Goloboff (1993), in which ischnothelines were found to be more closely related to Bipectina than euagrines. In our gene tree parsimony analyses, diplurid taxa form a monophyletic group, albeit with low support, and are included in a polytomy with Bipectina. This arrangement is somewhat similar to the findings in the ML and MP (with all spider taxa included) analyses of EF-1γ by Ayoub et al. (2007), which placed the diplurid taxa, *Euaegrus* and *Allothela*, as sister to each other and within bipectin spiders. The considerable incongruence in diplurid relationships among these various data sets clearly indicates the need for additional attention to this group. Equally challenging has been the determination of relationships among bipectin taxa. These relationships are largely unresolved based on our hemocyanin data set, although a theraphosid clade is consistently found sister to remaining bipectins. This bears some similarity to an early diverging theraphosid group in the ML analysis of EF-1γ (Ayoub et al., 2007). In addition, *Ancylotrypina* is consistently recovered as sister to the nemesiids, *Stanwellia* and *Acanthogonatus*, based on hemocyanins. In contrast, Hedin and Bond (2006) and Bond et al. (2012) recovered the “ancylotrypines” as the sister to theraphosids plus barychelids. While most nodes were unresolved among araneomorphs, hemocyanins support the traditional hypothesized sister relationship of *Hypochilus* to all other araneomorphs. This contrasts with EF-1γ, in which *Hypochilus* forms a monophyletic group with haplogyne spiders (Ayoub et al., 2007).

### 4.3. Molecular dating

Molecular clock analyses encompassing both mygalomorphs and araneomorphs have thus far been based on a single nuclear gene, EF-1γ (Ayoub and Hayashi, 2009; Ayoub et al., 2007). Our relaxed clock analyses based on hemocyanin sequences corroborate date estimates for certain nodes, and indicate nodal estimates that require further investigation. For most comparable nodes, 95% confidence intervals overlap between estimates based on hemocyanin and EF-1γ. Our divergence time estimate for the most recent common ancestor (MRCA) of Antrodiaetidae and Atypidae (node 21, Fig. 7; Table 3) at approximately 120 MYA is similar to that based on EF-1γ (140 MYA), being older than the minimum date of 96 MYA based on the two fossil Atypoidina spiders. The MRCA of diplurid taxa and more derived mygalomorphs (node 7) is estimated at 210 MYA, which is older than the estimates based on EF-1γ (190 MYA from MP, Ayoub and Hayashi, 2009; ~150 MYA based on ML, Ayoub et al., 2007). Divergence date estimates for the split between Araneomorphae and Mygalomorphae (node 1) are similar for hemocyanin (380 MYA) and EF-1γ (392 MYA).

A notable discrepancy between hemocyanin and EF-1γ occurs in the initial divergence dates among extant members within Mygalomorphae and Araneomorphae. Based on EF-1γ, Ayoub et al. (2007) found the age of the MRCA of Mygalomorphae (~295 MYA) to be much more recent than that of Araneomorphae (~375 MYA). The estimate of Ayoub et al. (2007) that the initial diversification of extant Araneomorphae preceded that of Mygalomorphae contrasts with hypotheses based strictly on fossils, which predict that the MRCA of extant mygalomorphs arose ~240 MYA, while the MRCA of extant araneomorphs arose more recently at ~180 MYA (Penny and Selden, 2007; Selden and Penny, 2010). The relative divergence dates of the two suborders based on hemocyanin are more similar to these fossil-based hypotheses in that the mygalomorph MRCA (node 5: ~330 MYA) occurred prior to the araneomorph MRCA (node 2: ~295 MYA), even though the calibration for orbicularians was ~30 million years older in our study compared to the calibration used for EF-1γ. However, confidence intervals for the initial mygalomorph and araneomorph divergences do overlap between EF-1γ and hemocyanin. Using multiple nuclear and mitochondrial loci, Dimitrov et al. (2011) estimated the initial araneomorph divergence date at ~325 MYA, which was also older than our hemocyanin
estimates. Mygalomorph divergences were not estimated in their study. The discrepancy in dates between hemocyanins, EF-1γ, and the multigene data set may be due to our sampling of paralogs, but may also reflect lineage specific rate variation among the different genes used in each study.

4.4. Conclusions

Hemocyanins play a crucial role in oxygen storage and transport in spiders, and this is reflected in the high sequence conservation in hemocyanin paralogs across distantly related species. However, hemocyanin gene family evolution has not always been static in spiders as reflected by the apparent loss of all paralog types but γ, followed by extensive duplication of this paralog in potentially all members of the RTA clade (Markl et al., 1986). Although causal mechanisms remain unknown, the diversification history of spider hemocyanins may correspond to the variation in respiratory demand across spider lineages associated with their different body sizes, activity levels, and respiratory structures. Our sampling strategy, while not comprehensive, does allow for many taxa to be screened for hemocyanin sequences, complementing studies using biochemical and cDNA methods that are typically restricted to smaller sample sizes.

Resolving ancient nodes in the spider phylogeny using molecular data has been a considerable challenge due to the few available molecular markers. The highly conserved hemocyanin sequences may help in this regard. Despite sampling across very distantly related spider species, the hemocyanin region we sequenced is intron free and alignment unambiguous over the 358 bp length. Yet, the sequences have sufficient variation (average ortholog group p-distances = 19–26%) to resolve deep relationships in the tree when concatenated. In spite of incongruence in relationships inferred from the different ortholog groups, phylogenetic analysis of the concatenated paralog data set produced a tree that is largely consistent with other recent molecular studies, particularly for mygalomorphs. Incongruence in relationships based on hemocyanins and other data sets indicates the need for sampling of additional molecular markers. Lastly, our molecular clock estimates based on hemocyanin sequences suggest that the initial divergences of mygalomorphs and araneomorphs occurred fairly close together in time (~300 MYA), with the initial mygalomorph divergence occurring only ~40 million years before that of araneomorphs.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

We thank Jason Bond, James Cokendolpher, Frederick Coyle, John Gatesy, Helene Lee, and Norman Platnick for their help with the collection of spiders and sequences. Comments of John Gatesy, Robert Meredith, and two anonymous reviewers helped to improve the manuscript. Financial support was provided by NSF DEB-0236020 (CYH) and NSF DEB-0108575 (MH and Jason Bond).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2013.04.037.

References
