

## In Depth Exploration of the Hemolymph of *Limulus polyphemus* via Combinatorial Peptide Ligand Libraries

Alfonsina D'Amato,<sup>†</sup> Angelo Cereda,<sup>‡</sup> Angela Bachi,<sup>†</sup> James C. Pierce,<sup>§</sup> and  
Pier Giorgio Righetti<sup>\*,‡</sup>

San Raffaele Scientific Institute, 20132 Milano, Italy, Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Via Mancinelli 7, 20131 Milano, Italy, and Department of Bioinformatics and Computer Science, University of the Sciences in Philadelphia, 600 South 43rd street, Philadelphia, Pennsylvania 19104-4495

Received March 5, 2010

The hemolymph of *Limulus polyphemus*, a very ancient marine arthropod dating back to ca. 440 million years, has been explored in depth via capture by combinatorial peptide ligand libraries. Whereas barely a dozen proteins had been known up to the present, we have increased this number by more than 1 order of magnitude, up to 160 unique gene products, identified via the dbEST\_limulus as well as via comparison with the other members of the Chelicerata subphylum to which *Limulus* belongs, namely, scorpions, ticks, mites, and spiders. Yet we have sequences of many other peptides, suggesting the presence of at least one more order of magnitude of species (1000 and more), that could not be identified as such sequences have no counterparts in present databases. This further reinforces the notion that these could be ancestral proteins, scarcely represented in present times. These data might represent the true birth of paleo-proteomics.

**Keywords:** peptide ligand libraries • low-abundance proteome • *Limulus polyphemus* • hemolymph • mass spectrometry • proteomics

### 1. Introduction

In recent times, paleo-genomics and paleo-proteomics were brought to the attention of the scientific community and of the general public by some remarkable reports. In the case of paleo-genomics, one wonders if the interest in this field was spurred by the famous novel *Jurassic Park* by Michael Crichton (and the subsequent, highly popular movie by Steven Spielberg), in which he proposed the recovery of dinosaur DNA from the alimentary tracts of hemophagous insects preserved for millions of years in amber.<sup>1</sup> The first scientific report appeared in 1994, when Woodward et al.<sup>2</sup> described the isolation of DNA fragments from a Late Cretaceous dinosaur bone preserved in bituminous strata, although the failure to authenticate this and other purported Mesozoic DNAs soon led to the consensus that DNA could not possibly survive much more than 100 000 years,<sup>3</sup> and even then, under the most extraordinary circumstances. Such circumstances might occur when fossils are preserved frozen in arctic ice masses. Thus, Willerslev et al.<sup>4</sup> recovered DNA frozen under two kilometers of glacial ice in Greenland dating back nearly a million years, while Bidle et al.<sup>5</sup> isolated microbial DNA from eight million year old Antarctic ice cores. Additionally, Salamon et al.<sup>6</sup> demonstrated that DNA

occluded within clusters of intergrown bone crystals was highly resistant to degradation.

In the case of paleo-proteomics, perhaps one of the reports that spurred much debate was the one by Asara et al.,<sup>7</sup> who claimed finding seven distinct collagen sequences by shotgun proteomics in a remarkably well-preserved 68-million-year-old fossilized dinosaur bone and computed a phylogenetic tree that placed *Tyrannosaurus rex* with birds. However, the discovery of intact protein in such an ancient sample was called into question on the grounds of plausibility<sup>8</sup> and inadequate statistical analysis.<sup>9</sup> This controversy was further revisited by Bern et al.,<sup>10</sup> who reanalyzed the original mass spectra yet again using different bioinformatics tools and statistical tests. Although they admit that the identification by Asara et al.<sup>7</sup> of bird-like collagen at the protein level is clearly significant, they also put forward a number of reservations, stating, for instance, that their "reanalysis shows a sample containing common laboratory contaminants and soil bacteria". And although they additionally remark that hemoglobin and collagen are plausible proteins to find in fossil bone, because they are two of the most abundant proteins in bone and bone marrow, they reinforce the notion that contamination remains a tricky and possibly unresolvable issue for this particular sample, to the point of suggesting that "perhaps a bird died on top of the *T. rex* excavation in the field and perhaps avian collagen from a cosmetic or medical product found its way into the *T. rex* sample". Perhaps a way to overcome both contaminants embedded in fossil samples and potential degradation of macromolecules within these specimens would be to analyze

\* To whom correspondence should be addressed. Fax: +39 02 23993080.  
E-mail: piergiorgio.righetti@polimi.it.

<sup>†</sup> San Raffaele Scientific Institute.

<sup>‡</sup> Politecnico di Milano.

<sup>§</sup> University of the Sciences in Philadelphia.

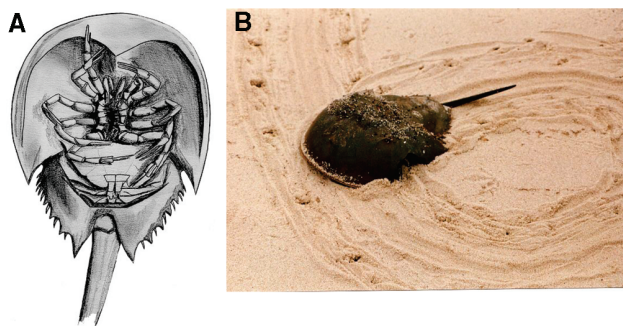


**Figure 1.** (A) Exoskeleton of *Limulus* (bottom side view). (B) Live *Limulus* caught while drawing the classical “O” of the Italian medieval painter Giotto on the sands of the Delaware Bay. Photos taken by author of this manuscript, James C. Pierce.

insects trapped within amber. Amber, the fossilized resins of trees, appears to be the most promising means enabling the preservation of proteins over millions of years. This resin is comprised largely of diterpenes, which rapidly polymerize, dehydrate the included specimen and, as they possess antimicrobial and anti-inflammatory properties, prevent decomposition and preserve the biological specimen intact for millions of years. Examples of exquisitely preserved biological specimens in amber abound in the literature.<sup>11,12</sup> As an example, transmission electron microscopy revealed that the morphology of cellular organelles such as nuclei, endoplasmic reticulum, ribosomes, and mitochondria was maintained in a 40 million year old fly imbibed in amber.<sup>13</sup> Yet, even in such well-shielded and preserved samples proteomic analysis is not immune from severe drawbacks. We recently extracted proteins from whole insects (bees, ants) entrapped in Dominican amber, 20–25 million years old, and subjected them to proteomic evaluation. The results were meagre: the SDS-extracted proteins were so highly cross-linked that they failed to penetrate even in large-pore SDS-PAGE gels and could hardly be digested by trypsin, producing peptides that had no counterpart in any database explored.<sup>14</sup>

A nice way to perform genuine paleo-proteomics, devoid of any possible pitfall and artifacts, would be to find a “living fossil” whose genomic and proteomic assets had been preserved unaltered and scarcely subjected to evolutionary pressure over the ages. An interesting example is the American horseshoe crab (*Limulus polyphemus*) (see Figure 1 for a rendering of this arthropod), which dates back to 440 million years ago, that is, almost at the beginning of the Phanerozoic period in earth.<sup>15</sup> Today there are only four species of horseshoe crabs reported, the American one (*Limulus polyphemus*) and three Asian species (*Tachypleus gigas*, *T. tridentatus* and *Carcinoscorpius rotundicauda*) that diverged from *Limulus* about 135 million years ago.<sup>16</sup> Curiously, very little is known on the proteomic asset of this marine animal. Just one protein has been extensively studied in detail, namely hemocyanin, a copper-containing respiratory protein found in the hemolymph of arthropods and molluscs.<sup>17</sup> Since these species do not confine their respiratory proteins into a cell, like the red blood cell (RBC) in humans, where hemoglobin is physically entrapped, hemocyanin, that is freely circulating in the lymph, is assembled as a multimeric protein reaching a size of 3.5 million Da. Yet in biology and the medical field the horseshoe crab is very well-known because of the “*Limulus* amebocyte lysate (LAL) test, used to detect bacterial contaminants (especially in

the form of lipopolysaccharides, LPS) in biologicals meant for human consumption, via a series of coagulation factors isolated from the haemocytes (also called amebocytes) circulating in the lymph. In the preparation of the LAL only the amebocytes are harvested and the lymph (serum) is discarded. Relatively little is known about the composition of the haemolymph, which in arthropods is the component of the circulatory system analogous to the fluids and cells making up the blood of higher animals. Only a dozen or so protein constituents of haemolymph are known at present, two of them, haemocyanin and complement C3 protein, well characterized due to their abundance, the others only known via their enzymatic activity.

A few years ago we reported a novel system for detecting low-abundance proteins in tissues and body fluids, based on a combinatorial peptide ligand library composed by several millions of hexapeptides bound to spheres constituted by an organic polymer (polymethacrylate).<sup>18</sup> Encouraged by the remarkable results obtained with a number of body fluids, such as human urines<sup>19</sup> and sera,<sup>20</sup> the cytoplasmic proteome of human RBCs,<sup>21</sup> and cerebrospinal fluid,<sup>22</sup> we have applied this technique to the in-depth exploration of the hemolymph of *Limulus*. Most encouraging and unique results have been obtained, as illustrated below, that suggest that paleo-proteomics should be coming of age.

## 2. Methods

**2.1. Materials, Equipments and Software.** The solid-phase combinatorial peptide library known under the trade name of ProteoMiner, as well as materials for electrophoresis such as gel plaques and reagents were from Bio-Rad Laboratories (Hercules, CA). *N*-Ethylmaleimide, urea, thiourea, 3-[3-cholamidopropyl dimethylammonio]-1-propanesulfonate (CHAPS), tris(2-carboxyethyl)phosphine hydrochloride, Bis-(2-hydroxyethyl)disulfide, isopropanol, acetonitrile, trifluoroacetic acid, and sodium dodecyl sulfate were all from Sigma-Aldrich (St Louis, Mo). Complete protease inhibitor cocktail tablets were from Roche Diagnostics, (Basel, Switzerland). Sequencing grade bovine trypsin was from Promega (Madison, WI). The capillary chromatographic system was EasyLC, from Proxeon Biosystem (Denmark). The capillary columns were homemade 10-cm reverse phase spraying fused silica (75  $\mu$ m i.d.  $\times$  10 cm), packed with 3- $\mu$ m ReproSil 100C18 (Dr. Maisch GmbH, Germany). The MS/MS analyses were performed by LTQ-Orbitrap mass spectrometer (ThermoScientific, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). The softwares for the data analysis were Mascot search engine (Matrix Science, London, U.K., version 2.2.06) and Scaffold (version Scaffold-01\_06\_07, Proteome Software Inc., Portland, OR).

**2.2. Hemolymph Preparation.** Adult female horseshoe crabs (*Limulus polyphemus*) were collected off of the southern coast of New Jersey and bled by nonlethal direct cardiac puncture. To prevent the blood cell hemocytes from coagulating, bleeding buffer (0.5 M NaCl, 0.01 M *N*-ethyl maleimide, and 1% Tween 20 v/v) was added immediately at the time of withdrawal of hemolymph (100 mL per 900 mL of hemolymph). The Roche protease inhibitor cocktail (containing 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), pepstatin A, E-64, bestatin, leupeptin, and aprotinin) was also added right away (1 mL per 1000 mL of hemolymph admixed with bleeding buffer) so as to prevent accidental proteolysis. The hemolymph was then centrifuged (5000 rpm for 10 min) so as to eliminate the hemocytes and obtain a particulate-free hemolymph. To

drastically reduce the high-abundance proteins (especially hemocyanin, present at a level of ca. 60%, similar to albumin concentration in human serum) three 15-mL aliquots of hemolymph were separately treated with 300  $\mu$ L of ProteoMiner beads at pH 4.0 (25 mM Na acetate in 50 mM KCl), at pH 7.0 (25 mM Na-phosphate in 50 mM KCl) and pH 9.5 (25 mM Tris-HCl in 50 mM KCl), as recommended by Fasoli et al.<sup>23</sup> Elution was implemented in 4% boiling SDS added with 25 mM DTT, as per Candiano et al.<sup>24</sup> For 2D map analyses, SDS was eliminated from the eluates in chloroform-methanol.

**2.3. Electrophoretic Analyses.** SDS-PAGE and 2D map analyses were performed exactly as described in Roux-Dalvai et al.<sup>21</sup>

**2.4. Protein Identification by nanoLC-MS/MS and Data Search.** The various sample lanes of SDS-PAGE gels were cut in 8 pieces of 0.5 to 1 cm along the migration path, and proteins were reduced by 10 mM DTT and alkylated by 55 mM iodoacetamide. The gel pieces were shrunk in acetonitrile and dried under vacuum; proteins were digested overnight with bovine trypsin as described elsewhere.<sup>25</sup> The tryptic mixtures were acidified with formic acid up to a final concentration of 1%. Five microliters of tryptic digest for each band were injected in a capillary chromatographic system. Peptide separations occurred on a reverse phase spraying fused silica capillary column. A gradient of eluents A (H<sub>2</sub>O with 2% v/v ACN, 0.1% v/v formic acid) and B (ACN with 2% v/v H<sub>2</sub>O with 0.1% v/v formic acid) was used to achieve separation, from: 8% B (at 0 min 0.2  $\mu$ L/min flow rate) to 50% B (at 80 min, 0.2  $\mu$ L/min flow rate). The LC system was connected to an LTQ-Orbitrap mass spectrometer. Full scan mass spectra were acquired in the LTQ Orbitrap mass spectrometer in the mass range  $m/z$  350 to 1500 Da with the resolution set to 60 000. The four most intense doubly and triply charged ions were automatically selected and fragmented in the ion trap. Target ions were selected at maximum two times for the MS/MS and were dynamically excluded for 60 s.<sup>25</sup> Data were searched by Mascot search engine against three databases: EST\_limulus (50 928 sequences; 954 406 residues); Uniprot\_chelicerata (version 15.13; 35 472 sequences; 8 793 815 residues) and Uniprot\_ixodes (version 15.13, 22980 sequences; 6 346 957 residues). We used tryptic cleavage constraints with maximum of 2 missed cleavages, cysteine alkylation and oxidation of methionine residues as variable modification. Peptide mass tolerance was set to 5 ppm and fragment mass tolerance to 0.6 Da. Scaffold was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm.<sup>26</sup> Protein identifications were accepted if they could be established at greater than 99.0% probability. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. The translated nucleotide sequences obtained by Mascot, consulting the EST database, were searched against nonredundant protein database by Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to find the homologous proteins. The Gene Ontology analysis was performed by using QuickGO interface (<http://www.ebi.ac.uk/QuickGO>).<sup>27</sup>

### 3. Results

Figure 2A shows a silver-stained 2D map of control, untreated hemolymph. Due to the overwhelming presence of

hemocyanin, seen as a string of bands covering the entire pI 3–10 range at around 73 kDa (and higher aggregates centered around pH 7 and giving a continuous smear up to 250 kDa) only about 200 hundred nonhemocyanin spots can be seen in the entire map. Conversely, the three eluates from ProteoMiner, admixed in a 1:1:1 and loaded in a 2D map, exhibit several hundred spots (890 via PD Quest count) with a distribution of spots covering especially the pH 5–8 region and ranging in size from ca. 8 kDa up to 240 kDa (Figure 2B).

For identifying all possible captured proteins, instead of eluting each individual spot in the 2D map, the control, untreated sample as well as the three ProteoMiner eluates at three different pH values (pH 4.0, 7.0 and 9.5) as per Fasoli et al.<sup>23</sup> were loaded onto a monodimensional SDS-PAGE gel (Figure 2C). Each Coomassie stained lane was cut into eight regions, trypsin digested and analyzed by nLC-MS/MS.<sup>25</sup> In Table 1 are reported a total of a 160 unique gene products. The homologous proteins found by Blast are also shown in Table 1. Surprisingly a consistent number of proteins was found homologous to ixodes, a genus of hard-bodied ticks. For this reason the MS/MS data were further analyzed by Mascot using ixodes and chelicerata databases. This allowed us to increase the number of *Limulus* proteins present in the hemolymph. Moreover some proteins, such as complement component 3 and apolipoprotein, were found more than one time due to their correspondence to different mRNA sequences. This increases the validity of identifications. However the list of identified proteins represents only a part of *Limulus*'s proteome. In the raw mass data, there is a large list of unmatched peptides that could substantially increase the identifications of more proteins. In detail the number of unmatched spectra is 31 252; considering the possibility that the same peptide ( $m/z$  value) can be sequenced at maximum two times (see Methods) even counting that the same peptide can be present as doubly or triply charged the resulting number of unmatched MS/MS is around 7500. On average, with a good LC-MS/MS run on the orbitrap we can achieve about 40–50% of identification rate that will lead to about 3–4000 more peptides identified on a known genome.<sup>28</sup>

Figure 3 gives overlapping Venn diagrams showing the capture and identifications of species in the three ProteoMiner eluates: it is seen that the best captures were those at pH 7.0 and 9.5. Upon eliminating the redundancies in the various eluates, a total of a 160 unique gene products could be listed in Table 1. Upon Gene Ontology analysis, a number of pathways could be categorized, as reported in Figure 4. The four major pathways are represented by oxygen transport activity (14%), transport processes (14%), oxidoreductase activity (11%), and metabolic processes (10%) (data referred to the pH 4.0 eluate; essentially identical values in all other eluates and control).

### 4. Discussion

Although not particularly emphasized in the previous sessions, the most arduous task turned out to be the identification of the proteins from the mass spectrometry data. When such data were confronted with *Limulus* databases, barely a handful of proteins could be identified, not many more than the 12 proteins already known from past literature. As luck goes, we became aware of a huge database that was published in January in a public repository on *Limulus* proteins (<http://www.ncbi.nlm.nih.gov/nucest?term=limulus>). It lists 8488 mRNA sequences obtained by homogenizing whole tissue of male

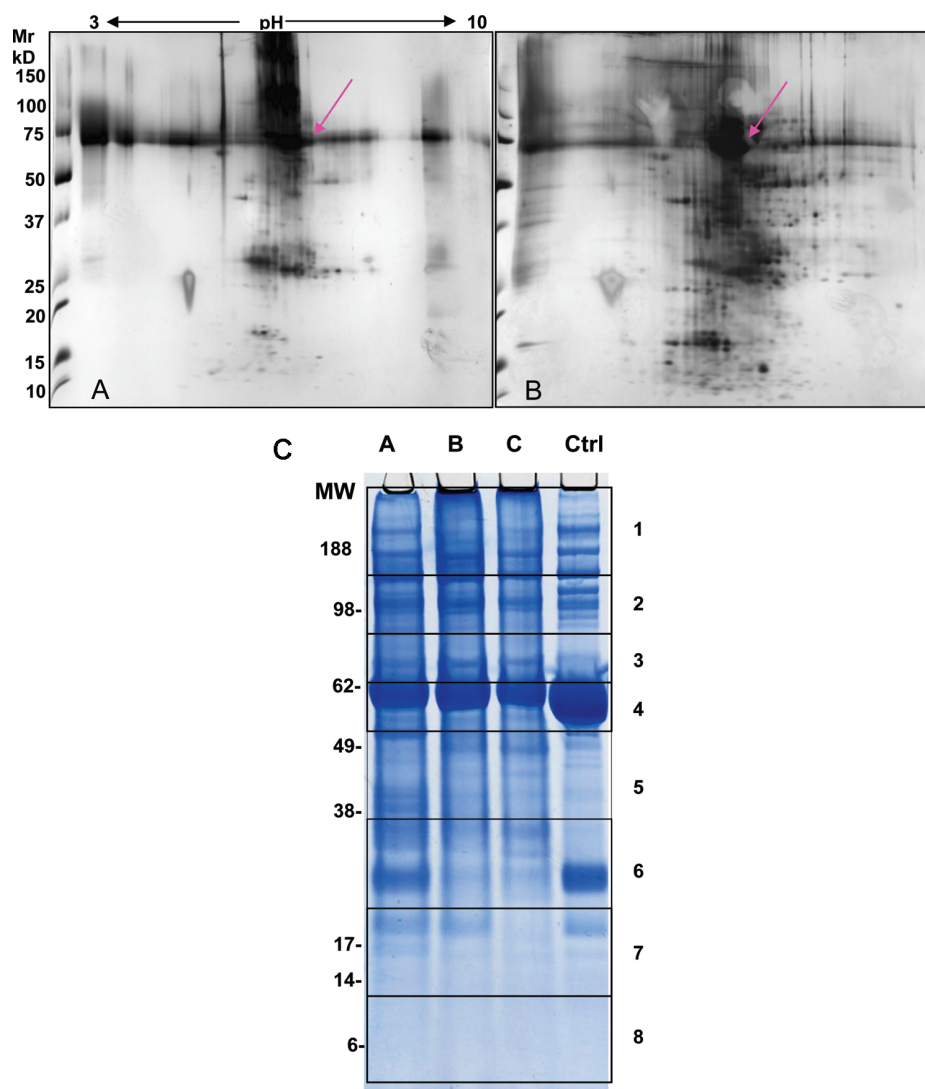


**Figure 2.** Two-dimensional maps of the hemolymph of *Limulus*. In both case the first dimension was an IPG pH 3–10 (non linear) strip, the second dimension an SDS gel in a linear porosity gradient of 8–18% polyacrylamide. Second dimension gel: 18 × 20 cm. Silver staining. Sample load: 600  $\mu$ g total protein. (A) Control lymph. (B) Combined eluates from ProteoMiner after capture at pH 4.0, 7.0, and 9.5. The arrows indicate the hemocyanin subunits (73 kDa) seen as multiple charge isoforms along the pH 3–10 gradient. (C) SDS-PAGE of the control, untreated sample as well as the three ProteoMiner eluates.

*Limulus polyphemus*, worked out by a group of German scientists in Berlin. When our peptides were searched against these sequence, a robust increment in novel protein identifications could be obtained (about 90), although a large number of identifications still regarded all the subunits constituting hemocyanin (i.e., the major lymph protein) and complement C3 protein, still the second most abundant protein in the lymph. Although this represented an increment of ca. 1 order of magnitude in respect to hemolymph constituents discovery, we were disappointed, as we had a vast body of sequenced peptides, suggesting the potential identification of >1000 proteins in this body fluid. It was brought to our attention that indeed horseshoe crabs are in the same subphylum Chelicerata, to which species that might appear to be totally unrelated (such as spiders, scorpions, ticks, and mites) belong. When our search was extended to embrace the databases of these additional species, all at once the number of identifications increased to a grand total of 160 unique gene products.

There are two important lessons that can be deduced from our data. First of all, whereas originally zoologists have assigned to this subphylum Chelicerata such apparently highly diverse species as mentioned above, purely on morphological examination, our proteomic data fully confirm their finding in an extraordinary way. For example, it can be appreciated from Table 1 that 67 identities were derived from the database of *Ixodes scapularis* (the tick) and others from scorpions, spiders, and mites. The other lesson we have learned is that indeed the proteins we could identify seem to be ancestral proteins, since in most cases the degree of homology with other species was between 30 and 40% and only in a few cases it was as high as 60–85%. This seems to suggest that such proteins have remained relatively unaltered over hundreds of millions of years, whereas in the other species in the class of Chelicerata they must have continuously evolved. For instance, it is surprising that, in the class of mites, more than 30 000 species have been reported, vs barely 4 different species of horseshoe

**Table 1.** One-Hundred Sixty Identified Proteins in Eluates A, B, C and Untreated Sample (ctrl) with Access Number, Molecular Mass, and Number of Peptides<sup>a</sup>

| dbEST_limulus database |        |                    |    |    | NCBI BLASTX                                                       |                                      |            |             |                     |                       |
|------------------------|--------|--------------------|----|----|-------------------------------------------------------------------|--------------------------------------|------------|-------------|---------------------|-----------------------|
| accession number       | MW     | number of peptides |    |    | protein name                                                      | taxonomy                             | % homology | Blast score | e-value             | NCBI accession number |
|                        |        | A                  | B  | C  |                                                                   |                                      |            |             |                     |                       |
| gi114106953            | 45 kDa | 2                  | 0  | 0  | MUC5AC protein                                                    | <i>Ixodes scapularis</i>             | 48         | 50-80       | $4 \times 10^{-15}$ | XM_002407179.1        |
| gi114106961            | 45 kDa | 0                  | 2  | 3  | endotoxin-binding protein                                         | <i>Limulus polyphemus</i>            | 96         | >200        | $4 \times 10^{-58}$ | M65017.1              |
| gi283495093            | 55 kDa | 2                  | 2  | 2  | conserved hypothetical protein                                    | <i>Ixodes scapularis</i>             | 75         | >200        | $3 \times 10^{-52}$ | XM_002406239.1        |
| gi283495119            | 50 kDa | 2                  | 0  | 1  | carcinolectin-5C                                                  | <i>Carcinoscorpius rotundicauda</i>  | 50         | 80-200      | $8 \times 10^{-46}$ | DQ250746.1            |
| gi283495152            | 49 kDa | 9                  | 10 | 10 | apolipoprotein                                                    | <i>Ixodes scapularis</i>             | 41         | 80-200      | $2 \times 10^{-19}$ | XM_002401724.1        |
| gi283495250            | 55 kDa | 2                  | 2  | 6  | muscular protein 20                                               | <i>Zophosis dilatata</i>             | 66         | 80-200      | $8 \times 10^{-38}$ | FN545124.1            |
| gi283495327            | 48 kDa | 0                  | 2  | 0  | similar to adenosylhomocysteinase                                 | <i>Tribolium castaneum</i>           | 79         | >200        | $7 \times 10^{-54}$ | XM_963273.2           |
| gi283495352            | 51 kDa | 0                  | 2  | 0  | malate dehydrogenase                                              | <i>Pedicularius humanus corporis</i> | 79         | 80-200      | $3 \times 10^{-43}$ | XM_002424763.1        |
| gi283495374            | 50 kDa | 4                  | 8  | 4  | glutathione peroxidase                                            | <i>Ixodes ricinus</i>                | 70         | 80-200      | $1 \times 10^{-43}$ | EF21348.1             |
| gi283495527            | 41 kDa | 4                  | 3  | 5  | aldehyde dehydrogenase                                            | <i>Ixodes scapularis</i>             | 70         | >200        | $2 \times 10^{-45}$ | XM_002415019.1        |
| gi283495565            | 52 kDa | 5                  | 3  | 4  | hypothetical protein                                              | <i>Branchiostoma floridae</i>        | 37         | 50-80       | $1 \times 10^{-12}$ | XM_002593527.1        |
| gi283495592            | 54 kDa | 14                 | 4  | 11 | hypothetical protein                                              | <i>Ixodes scapularis</i>             | 31         | 50-80       | 0.0000004           | XM_002401721.1        |
| gi283495611            | 54 kDa | 0                  | 12 | 0  | sapoin                                                            | <i>Ixodes scapularis</i>             | 37         | 80-200      | $7 \times 10^{-25}$ | XM_002412013.1        |
| gi283495628            | 55 kDa | 0                  | 4  | 0  | acid methyltransferase                                            | <i>Ixodes scapularis</i>             | 36         | 50-80       | $6 \times 10^{-12}$ | XM_002401315.1        |
| gi283495630            | 45 kDa | 0                  | 1  | 1  | limulus intracellular coagulation                                 | <i>Tachypleus tridentatus</i>        | 50         | 50-80       | $4 \times 10^{-10}$ | D32211.1              |
| gi283495754            | 52 kDa | 3                  | 1  | 2  | inhibitor type 2 precursor                                        | <i>Drosophila melanogaster</i>       | 52         | 80-200      | $1 \times 10^{-42}$ | AAB20908.1            |
| gi283495762            | 51 kDa | 2                  | 3  | 2  | glutathione S-transferase 27                                      | <i>Ixodes scapularis</i>             | 82         | 80-200      | $6 \times 10^{-46}$ | XM_002408335.1        |
| gi283495787            | 53 kDa | 0                  | 2  | 3  | enolase, putative                                                 | <i>Limulus polyphemus</i>            | 100        | >200        | $4 \times 10^{-64}$ | X04424.1              |
| gi283495850            | 60 kDa | 2                  | 2  | 2  | coagulogen                                                        |                                      |            |             |                     |                       |
| gi283495855            | 60 kDa | 0                  | 0  | 3  | NO                                                                |                                      |            |             |                     |                       |
| gi283495853            | 47 kDa | 0                  | 2  | 1  | hypothetical protein                                              | <i>Ixodes scapularis</i>             | 22         | 40-50       | 0.21                | XM_002411134.1        |
| gi283495933            | 52 kDa | 0                  | 2  | 1  | fructose 1,6-bisphosphate aldol                                   | <i>Ixodes scapularis</i>             | 76         | 80-200      | $2 \times 10^{-49}$ | XM_002411723.1        |
| gi2834959975           | 52 kDa | 0                  | 2  | 2  | nucleotide excision repair factor                                 | <i>Ixodes scapularis</i>             | 47         | 80-200      | $2 \times 10^{-33}$ | XM_002412066.1        |
| gi283500021            | 46 kDa | 0                  | 2  | 0  | NEF2, RAD23 component                                             |                                      |            |             |                     |                       |
|                        |        |                    |    |    | alpha-tubulin mRNA, clone                                         |                                      |            |             |                     |                       |
|                        |        |                    |    |    | Ptalpha2                                                          | <i>Pelvetia fastigiata</i>           | 72         | 80-200      | $5 \times 10^{-24}$ | gbU58642.1            |
| gi283500358            | 41 kDa | 0                  | 2  | 0  | ribosomal protein P2                                              | <i>Ictalurus punctatus</i>           | 75         | 80-200      | $4 \times 10^{-20}$ | AF401553.1            |
| gi283500525            | 12 kDa | 1                  | 0  | 0  | ATP synthase F0 subunit 6                                         | <i>Limulus polyphemus</i>            | 84         | 80-200      | $1 \times 10^{-30}$ | NC_003057.1           |
| gi283500577            | 54 kDa | 0                  | 0  | 5  | serum amyloid A protein                                           | <i>Ixodes scapularis</i>             | 46         | 80-200      | $2 \times 10^{-21}$ | XM_002407273.1        |
| gi283500588            | 53 kDa | 3                  | 2  | 2  | ubiquitin (ribosomal protein L40)                                 | <i>Schistosoma mansoni</i>           | 100        | >200        | $3 \times 10^{-93}$ | XM_002575944.1        |
| gi283500594            | 55 kDa | 0                  | 2  | 0  | similar to saposin isoform 1                                      | <i>Tribolium castaneum</i>           | 32         | 80-200      | $4 \times 10^{-15}$ | XM_961759.2           |
| gi283500631            | 53 kDa | 0                  | 2  | 2  | translation elongation factor EF-1                                | <i>Ixodes scapularis</i>             | 95         | 80-200      | $8 \times 10^{-47}$ | XM_002411102.1        |
|                        |        |                    |    |    | alpha/Tu                                                          |                                      |            |             |                     |                       |
| gi283500653            | 53 kDa | 0                  | 0  | 2  | ADP ribosylation factor 79F                                       | <i>Tribolium castaneum</i>           | 96         | >200        | $4 \times 10^{-69}$ | XM_963294.2           |
| gi283500743            | 50 kDa | 2                  | 0  | 2  | 1A family penicillin-binding protein                              | <i>Ruegeria pomeroyi</i> DSS3        | 29         | <40         | 1.3                 | NC_003911.1           |
| gi283500755            | 51 kDa | 0                  | 3  | 3  | beta tubulin                                                      | <i>Ixodes scapularis</i>             | 99         | >200        | $1 \times 10^{-80}$ | XM_002406620.1        |
| gi283500908            | 17 kDa | 2                  | 2  | 0  | NO                                                                |                                      |            |             |                     |                       |
| gi283500914            | 54 kDa | 0                  | 2  | 2  | Mapmodulin                                                        | <i>Ixodes scapularis</i>             | 87         | >200        | $7 \times 10^{-58}$ | XM_002407010.1        |
| gi283500995            | 56 kDa | 1                  | 2  | 1  | carcinolectin5b-9                                                 | <i>Carcinoscorpius rotundicauda</i>  | 49         | 80-200      | $3 \times 10^{-29}$ | DQ841203.1            |
| gi283501103            | 58 kDa | 0                  | 7  | 11 | actin                                                             | <i>Hypochitlus thorelli</i>          | 98         | >200        | $4 \times 10^{-74}$ | EU293222.1            |
| gi283501111            | 54 kDa | 0                  | 2  | 0  | proteasome alpha subunit                                          | <i>Ixodes scapularis</i>             | 82         | >200        | $4 \times 10^{-70}$ | XM_002435334.1        |
| gi283501118            | 55 kDa | 2                  | 0  | 0  | calponin                                                          | <i>Ixodes scapularis</i>             | 68         | >200        | $4 \times 10^{-61}$ | XM_002401975.1        |
| gi283501140            | 52 kDa | 0                  | 0  | 2  | Tachylectin-P                                                     | <i>Tachypleus tridentatus</i>        | 80         | >200        | $7 \times 10^{-91}$ | AB028144.1            |
| gi283501188            | 51 kDa | 1                  | 0  | 0  | origin recognition complex, second largest subunit ORC2, putative | <i>Ixodes scapularis</i>             | 26         | 40-50       | 0.17                | XM_002435030.1        |
| gi283501225            | 49 kDa | 2                  | 3  | 2  | DNA-binding protein, putative                                     | <i>Ixodes scapularis</i>             | 73         | >200        | $2 \times 10^{-72}$ | XM_002403547.1        |
| gi283501255            | 52 kDa | 0                  | 0  | 1  | neurogenic locus notch                                            | <i>Ixodes scapularis</i>             | 54         | 80-200      | $7 \times 10^{-43}$ | XM_002411393.1        |
| gi283501274            | 56 kDa | 2                  | 2  | 2  | alpha-2-macroglobulin                                             | <i>Limulus sp.</i>                   | 96         | >200        | $6 \times 10^{-75}$ | D83196.1              |
| gi283501390            | 56 kDa | 0                  | 0  | 2  | hypothetical protein                                              | <i>Ixodes scapularis</i>             | 53         | 80-200      | $7 \times 10^{-29}$ | XM_002416298.1        |
|                        |        |                    |    |    | Iscw_ISCW023508                                                   |                                      |            |             |                     |                       |
| gi283511130            | 52 kDa | 0                  | 3  | 2  | Histone H3.3                                                      | <i>Salmo salar</i>                   | 99         | >200        | $6 \times 10^{-57}$ | BT057313.1            |
| gi283511135            | 35 kDa | 0                  | 0  | 0  | octin2.2b                                                         | <i>Carcinoscorpius rotundicauda</i>  | 54         | 80-200      | $1 \times 10^{-19}$ | DQ648079.1            |
| gi283512302            | 49 kDa | 0                  | 1  | 0  | Actin-related protein 2/3 complex subunit 4                       | <i>Anoplopona fimbria</i>            | 90         | 80-200      | $6 \times 10^{-43}$ | BT083058.1            |
| gi283512334            | 48 kDa | 0                  | 2  | 2  | Chain A, Crystal Structure Of Creatine-                           | <i>Limulus polyphemus</i>            | 96         | >200        | $8 \times 10^{-95}$ | P51541                |

Table 1. Continued

| dbEST_limulus database |        |                    |    |    | NCBI BLASTX                                                             |                                      |            |             |                        |                       |
|------------------------|--------|--------------------|----|----|-------------------------------------------------------------------------|--------------------------------------|------------|-------------|------------------------|-----------------------|
| accession number       | MW     | number of peptides |    |    | protein name                                                            | taxonomy                             | % homology | Blast score | e-value                | NCBI accession number |
|                        |        | A                  | B  | C  |                                                                         |                                      |            |             |                        |                       |
| gi 283512358           | 42 kDa | 0                  | 0  | 0  | 3 viral A-type inclusion protein                                        | <i>Trichomonas vaginalis</i> G3      | 26         | <40         | 2.3                    | XM_001311356.1        |
| gi 283512604           | 46 kDa | 3                  | 2  | 0  | 4 L-cystatin                                                            | <i>Tachypleus tridentatus</i>        | 57         | 80–200      | 1 × 10 <sup>-16</sup>  | Q7M429                |
| gi 283512836           | 52 kDa | 0                  | 6  | 3  | 0 Arginine kinase                                                       | <i>Limulus polyphemus</i>            | 100        | >200        | 6 × 10 <sup>-58</sup>  | P51541                |
| gi 283512864           | 49 kDa | 0                  | 1  | 0  | 0 muscle LIM protein isoform A                                          | <i>Ixodes scapularis</i>             | 84         | 80–200      | 1 × 10 <sup>-20</sup>  | XM_002434020.1        |
| gi 283512904           | 56 kDa | 1                  | 2  | 0  | 3 cyclophilin A                                                         | <i>Ixodes scapularis</i>             | 84         | >200        | 6 × 10 <sup>-78</sup>  | XM_002407873.1        |
| gi 283513061           | 56 kDa | 0                  | 2  | 0  | 2 gene product transcript GI10977-RA                                    | <i>Drosophila mojavensis</i>         | 29         | 40–50       | 0.001                  | XM_002011614.1        |
| gi 283514106           | 58 kDa | 10                 | 8  | 6  | 4 selenium dependent salivary glutathione peroxidase                    | <i>Ixodes scapularis</i>             | 55         | 80–200      | 5 × 10 <sup>-26</sup>  | DQ066177.1            |
| gi 283514182           | 51 kDa | 3                  | 2  | 2  | 0 Chain A, Hydroxo Bridge Met Form Hemocyanin                           | <i>Limulus polyphemus</i>            | 98         | 80–200      | 9 × 10 <sup>-45</sup>  | AM260213.1            |
| gi 283515916           | 50 kDa | 2                  | 0  | 0  | 1 hypothetical protein EBL_27100                                        | <i>Enterocytozoan bienewsi</i> H348  | 32         | <40         | 0.14                   | XM_002649740.1        |
| gi 283519488           | 56 kDa | 2                  | 2  | 2  | 0 similar to CG6330-PA, isoform A                                       | <i>Apis mellifera</i>                | 60         | 80–200      | 2 × 10 <sup>-35</sup>  | XM_395069.2           |
| gi 283522680           | 51 kDa | 1                  | 5  | 6  | 1 actin-5                                                               | <i>Limulus polyphemus</i>            | 100        | >200        | 1 × 10 <sup>-79</sup>  | P41339                |
| gi 283522719           | 49 kDa | 2                  | 2  | 0  | 0 cholinergic receptor nicotinic alpha polypeptide 1                    | <i>Takifugu rubripes</i>             | 31         | 80–200      | 2 × 10 <sup>-22</sup>  | DQ481668.1            |
| gi 283525138           | 46 kDa | 0                  | 6  | 4  | 5 keratinoxin                                                           | <i>Tachypleus tridentatus</i>        | 75         | >200        | 1 × 10 <sup>-50</sup>  | AB201713.1            |
| gi 283525272           | 48 kDa | 2                  | 5  | 4  | 0 fasciclin, putative                                                   | <i>Ixodes scapularis</i>             | 38         | 50–80       | 3 × 10 <sup>-13</sup>  | XM_002409944.1        |
| gi 283525425           | 47 kDa | 3                  | 3  | 3  | 5 SAP-like pentraxin                                                    | <i>Limulus polyphemus</i>            | 99         | >200        | 3 × 10 <sup>-84</sup>  | AY066022.1            |
| gi 283525938           | 48 kDa | 0                  | 4  | 0  | 4 glyoxylate/hydroxypyruvate reductase                                  | <i>Ixodes scapularis</i>             | 61         | 80–200      | 8 × 10 <sup>-48</sup>  | XM_002407106.1        |
| gi 283539007           | 57 kDa | 7                  | 4  | 3  | 6 c reactive protein precursor similar to 14–3–3 CGI7870-PA, isoform A  | <i>Limulus polyphemus</i>            | 98         | >200        | 8 × 10 <sup>-104</sup> | P06205                |
| gi 283539103           | 50 kDa | 0                  | 2  | 0  | 0 Isoform A                                                             | <i>Apis mellifera</i>                | 93         | 80–200      | 6 × 10 <sup>-41</sup>  | XM_623180.2           |
| gi 283539209           | 54 kDa | 1                  | 0  | 2  | 1 GTP-binding protein, putative                                         | <i>Ixodes scapularis</i>             | 67         | 80–200      | 3 × 10 <sup>-41</sup>  | XM_002411991.1        |
| gi 283539310           | 53 kDa | 4                  | 0  | 0  | 0 serpin-8 precursor                                                    | <i>Ixodes scapularis</i>             | 41         | 80–200      | 3 × 10 <sup>-27</sup>  | XM_002415263.1        |
| gi 283539314           | 55 kDa | 0                  | 2  | 0  | 2 histone H2B                                                           | <i>Ixodes scapularis</i>             | 91         | 80–200      | 3 × 10 <sup>-43</sup>  | XM_002402577.1        |
| gi 283539467           | 53 kDa | 19                 | 17 | 8  | 19 secreted salivary gland peptide                                      | <i>Ixodes scapularis</i>             | 43         | 80–200      | 8 × 10 <sup>-39</sup>  | XM_002406216.1        |
| gi 283539482           | 50 kDa | 0                  | 2  | 0  | 0 DNA replication factor/protein phosphatase inhibitor SET/SPR-2        | <i>Ixodes scapularis</i>             | 91         | 80–200      | 2 × 10 <sup>-41</sup>  | XM_002411657.1        |
| gi 283541469           | 52 kDa | 3                  | 0  | 0  | 2 acetylcholinesterase                                                  | <i>Ixodes scapularis</i>             | 57         | >200        | 3 × 10 <sup>-53</sup>  | XM_002413167.1        |
| gi 283541621           | 53 kDa | 0                  | 5  | 6  | 0 SJHGC08169 protein                                                    | <i>Schistosoma japonicum</i>         | 25         | <40         | 1.5                    | AY810201.1            |
| gi 283541679           | 55 kDa | 2                  | 3  | 2  | 0 elongation factor-2                                                   | <i>Limulus polyphemus</i>            | 100        | >200        | 1 × 10 <sup>-83</sup>  | AY305496.1            |
| gi 283541875           | 58 kDa | 0                  | 4  | 3  | 0 predicted protein                                                     | <i>Naegleria gruberi</i>             | 22         | 50–80       | 0.000008               | GG738845.1            |
| gi 283541987           | 53 kDa | 0                  | 1  | 2  | 0 GI10977                                                               | <i>Drosophila mojavensis</i>         | 29         | 40–50       | 0.001                  | XM_002011614.1        |
| gi 283542021           | 51 kDa | 0                  | 2  | 0  | 0 peroxiredoxin                                                         | <i>Femmeropenaeus indicus</i>        | 83         | >200        | 1 × 10 <sup>-61</sup>  | GQ161914              |
| gi 283543137           | 56 kDa | 0                  | 2  | 1  | 0 complement component factor B/C2 similar to Chain A, The Structure Of | <i>Branchiostoma belcheri</i>        | 31         | 50–80       | 0.0000002              | EU316215.1            |
| gi 283543154           | 53 kDa | 0                  | 2  | 0  | 0 Alpha-N-Acetylgalactosaminidase                                       | <i>Strongylocentrotus purpuratus</i> | 71         | >200        | 1 × 10 <sup>-73</sup>  | XM_787549.2           |
| gi 283543216           | 54 kDa | 5                  | 4  | 5  | 4 selenoprotein P                                                       | <i>Saccoglossus kowalevskii</i>      | 50         | 50–80       | 5 × 10 <sup>-12</sup>  | GU224228.1            |
| gi 283543382           | 59 kDa | 5                  | 0  | 0  | 2 secreted putative protein                                             | <i>Ixodes scapularis</i>             | 49         | >200        | 9 × 10 <sup>-51</sup>  | XM_002435904.1        |
| gi 283543443           | 51 kDa | 18                 | 16 | 20 | 15 complement component 3-like protein                                  | <i>Carcinoscorpius rotundicauda</i>  | 89         | 80–200      | 2 × 10 <sup>-45</sup>  | AF517564.1            |
| gi 283543489           | 58 kDa | 2                  | 0  | 2  | 3 NO                                                                    |                                      |            |             |                        |                       |
| gi 283543670           | 50 kDa | 0                  | 3  | 0  | 0 sorbitol dehydrogenase                                                | <i>Ixodes scapularis</i>             | 72         | >200        | 6 × 10 <sup>-56</sup>  | XM_002412565.1        |
| gi 283543751           | 54 kDa | 0                  | 1  | 0  | 0 alpha-L-fucosidase                                                    | <i>Ixodes scapularis</i>             | 67         | >200        | 2 × 10 <sup>-74</sup>  | XM_002131448.1        |
| gi 283543849           | 46 kDa | 2                  | 2  | 0  | 0 transcriptional regulator                                             | <i>Enterococcus faecium</i> Com12    | 33         | <40         | 0.29                   | NZ_GG570311.1         |

Table 1. Continued

| accession number | MW      | number of peptides        |    |    |      | protein name                                                   | taxonomy                            |
|------------------|---------|---------------------------|----|----|------|----------------------------------------------------------------|-------------------------------------|
|                  |         | A                         | B  | C  | ctrl |                                                                |                                     |
|                  |         | Uniprot Chelicer database |    |    |      |                                                                |                                     |
| A1KYP6           | 24 kDa  | 0                         | 2  | 6  | 3    | Plasma carcinolectin CL5A1                                     | <i>Carcinoscorpius rotundicauda</i> |
| A1KYQ1           | 31 kDa  | 0                         | 0  | 2  | 0    | Plasma carcinolectin CL5B1 (Fragment)                          | <i>Carcinoscorpius rotundicauda</i> |
| A1XIV1           | 72 kDa  | 3                         | 2  | 3  | 4    | Hemocyanin subunit I                                           | <i>Carcinoscorpius rotundicauda</i> |
| A1XIV2           | 73 kDa  | 2                         | 0  | 2  | 0    | Hemocyanin subunit II                                          | <i>Carcinoscorpius rotundicauda</i> |
| A1XIV3           | 73 kDa  | 4                         | 0  | 0  | 1    | Hemocyanin subunit IIIa                                        | <i>Carcinoscorpius rotundicauda</i> |
| A1XIV4           | 73 kDa  | 26                        | 22 | 24 | 24   | Hemocyanin subunit IIIb                                        | <i>Carcinoscorpius rotundicauda</i> |
| A1XIV6           | 74 kDa  | 12                        | 10 | 10 | 12   | Hemocyanin subunit V                                           | <i>Carcinoscorpius rotundicauda</i> |
| A2AX56           | 73 kDa  | 2                         | 2  | 2  | 3    | Hemocyanin subunit II                                          | <i>Limulus polyphemus</i>           |
| A2AX57           | 72 kDa  | 61                        | 42 | 48 | 61   | Hemocyanin subunit IIIa                                        | <i>Limulus polyphemus</i>           |
| A2AX58           | 73 kDa  | 67                        | 50 | 59 | 68   | Hemocyanin subunit IV                                          | <i>Limulus polyphemus</i>           |
| A2AX59           | 74 kDa  | 62                        | 47 | 52 | 57   | Hemocyanin subunit VI                                          | <i>Limulus polyphemus</i>           |
| A3RIQ3           | 15 kDa  | 0                         | 3  | 2  | 0    | Histone H3 (Fragment)                                          | <i>Portia labiata</i>               |
| A4UTU3           | 42 kDa  | 0                         | 0  | 3  | 0    | Beta-actin                                                     | <i>Dermacentor variabilis</i>       |
| A5Z1D9           | 33 kDa  | 0                         | 2  | 0  | 0    | Putative uncharacterized protein                               | <i>Haemaphysalis qinghaiensis</i>   |
| A6N9Z0           | 18 kDa  | 3                         | 3  | 2  | 2    | Ubiquitin/40S ribosomal protein S27a                           | <i>Ornithodoros parkeri</i>         |
| A9X251           | 32 kDa  | 2                         | 2  | 2  | 2    | Oc1n 1.1                                                       | <i>Carcinoscorpius rotundicauda</i> |
| A9X6Z9           | 36 kDa  | 0                         | 0  | 2  | 0    | Putative C-1-tetrahydrofolate synthetase                       | <i>Carcinoscorpius rotundicauda</i> |
| A9XYM6           | 22 kDa  | 0                         | 0  | 2  | 0    | Carcinolectin5b-6                                              | <i>Limulus polyphemus</i>           |
| B2YGD6           | 22 kDa  | 0                         | 0  | 3  | 0    | Actin (Fragment)                                               | <i>Goleba lyra</i>                  |
| B2ZW74           | 20 kDa  | 2                         | 0  | 0  | 2    | Cyclophilin A                                                  | <i>Haemaphysalis longicornis</i>    |
| B6ZH52           | 196 kDa | 53                        | 46 | 56 | 43   | Complement component 3                                         | <i>Tachyplesus tridentatus</i>      |
| B7P1K3           | 146 kDa | 1                         | 0  | 0  | 0    | SMC protein, putative (Fragment)                               | <i>Isodex scapularis</i>            |
| B7PA95           | 50 kDa  | 0                         | 3  | 4  | 0    | Beta tubulin                                                   | <i>Isodex scapularis</i>            |
| B7PB81           | 27 kDa  | 0                         | 2  | 0  | 0    | Putative uncharacterized protein                               | <i>Isodex scapularis</i>            |
| B7PCG8           | 55 kDa  | 0                         | 3  | 2  | 0    | Alpha tubulin, putative                                        | <i>Isodex scapularis</i>            |
| B7PHT2           | 15 kDa  | 3                         | 2  | 0  | 2    | Histone H2A                                                    | <i>Isodex scapularis</i>            |
| B7PRG2           | 35 kDa  | 0                         | 2  | 0  | 0    | 60S acidic ribosomal protein P0                                | <i>Isodex scapularis</i>            |
| B7Q349           | 51 kDa  | 0                         | 4  | 3  | 0    | Elongation factor 1-alpha                                      | <i>Isodex scapularis</i>            |
| B7Q573           | 28 kDa  | 0                         | 2  | 0  | 0    | DNA replication factor/protein phosphatase inhibitor SET/SPR-2 | <i>Isodex scapularis</i>            |
| B7Q514           | 30 kDa  | 0                         | 3  | 0  | 0    | Multifunctional chaperone, putative                            | <i>Isodex scapularis</i>            |
| B7QIS6           | 54 kDa  | 49                        | 2  | 0  | 2    | Aldehyde dehydrogenase, putative                               | <i>Isodex scapularis</i>            |
| O01717           | 168 kDa | 11                        | 38 | 32 | 78   | Alpha-2-macroglobulin                                          | <i>Limulus sp.</i>                  |
| P02241           | 72 kDa  | 6                         | 0  | 1  | 2    | Hemocyanin D chain                                             | <i>Aphonopelma</i>                  |
| P03998           | 22 kDa  | 0                         | 3  | 3  | 7    | Coagulogen                                                     | <i>Limulus polyphemus</i>           |
| P04253           | 73 kDa  | 2                         | 36 | 46 | 42   | Hemocyanin II                                                  | <i>Limulus polyphemus</i>           |
| P06205           | 27 kDa  | 8                         | 7  | 7  | 11   | C-reactive protein 1.1                                         | <i>Limulus polyphemus</i>           |
| P06206           | 27 kDa  | 2                         | 0  | 2  | 4    | C-reactive protein 1.4                                         | <i>Limulus polyphemus</i>           |
| P07086           | 12 kDa  | 0                         | 1  | 0  | 0    | Antilipopolysaccharide factor                                  | <i>Limulus polyphemus</i>           |
| P28175           | 112 kDa | 2                         | 0  | 3  | 4    | Limulus clotting factor C                                      | <i>Tachyplesus tridentatus</i>      |
| P41339           | 42 kDa  | 2                         | 9  | 17 | 5    | Actin, acrosomal process isoform                               | <i>Limulus polyphemus</i>           |
| P51541           | 40 kDa  | 2                         | 9  | 5  | 0    | Arginine kinase                                                | <i>Limulus polyphemus</i>           |
| Q25387           | 15 kDa  | 0                         | 2  | 2  | 3    | Endotoxin-binding protein-protease inhibitor                   | <i>Limulus polyphemus</i>           |
| Q27085           | 48 kDa  | 2                         | 2  | 0  | 0    | Intracellular coagulation inhibitor                            | <i>Limulus polyphemus</i>           |
| Q2T530           | 29 kDa  | 0                         | 2  | 2  | 2    | Galactose-binding protein                                      | <i>Tachyplesus tridentatus</i>      |
| Q4PM69           | 11 kDa  | 2                         | 5  | 3  | 2    | Histone H4                                                     | <i>Carcinoscorpius rotundicauda</i> |
| Q6QWN8           | 32 kDa  | 0                         | 3  | 0  | 0    | Glyceraldehyde-3-phosphate dehydrogenase                       | <i>Isodex scapularis</i>            |
| Q6QWP9           | 40 kDa  | 2                         | 3  | 3  | 2    | Enolase (Fragment)                                             | <i>Limulus polyphemus</i>           |
| Q6XP51           | 24 kDa  | 0                         | 2  | 0  | 0    | Rab11-1a                                                       | <i>Limulus polyphemus</i>           |
| Q7M4H0           | 3 kDa   | 2                         | 0  | 2  | 2    | Hemocyanin subunit IIIb (Fragment)                             | <i>Limulus polyphemus</i>           |
| Q7M4H2           | 3 kDa   | 3                         | 3  | 2  | 3    | Hemocyanin subunit I (Fragment)                                | <i>Limulus polyphemus</i>           |
| Q8WQK3           | 26 kDa  | 2                         | 3  | 3  | 6    | SAP-like pentraxin                                             | <i>Limulus polyphemus</i>           |
| Q94823           | 46 kDa  | 2                         | 2  | 3  | 2    | Intracellular coagulation inhibitor type3                      | <i>Tachyplesus tridentatus</i>      |

Table 1. Continued

| Uniprot Ixodes database |         |                    |    |    |      |                                          |                          |
|-------------------------|---------|--------------------|----|----|------|------------------------------------------|--------------------------|
| accession number        | MW      | number of peptides |    |    |      | taxonomy                                 |                          |
|                         |         | A                  | B  | C  | Ctrl |                                          |                          |
| B7P1K3                  | 146 kDa | 2                  | 1  | 0  | 0    | SMC protein, putative (Fragment)         | <i>Ixodes scapularis</i> |
| B7P3N5                  | 290 kDa | 1                  | 0  | 0  | 0    | Ran-binding protein (RanBP), putative    | <i>Ixodes scapularis</i> |
| B7P438                  | 33 kDa  | 1                  | 0  | 0  | 0    | Calcium-binding protein, putative        | <i>Ixodes scapularis</i> |
| B7P7T4                  | 15 kDa  | 1                  | 3  | 2  | 2    | Histone H3                               | <i>Ixodes scapularis</i> |
| B7PBG2                  | 38 kDa  | 3                  | 11 | 14 | 5    | Actin, putative (Fragment)               | <i>Ixodes scapularis</i> |
| B7PD73                  | 30 kDa  | 2                  | 2  | 2  | 0    | Alpha-L-fucosidase, putative             | <i>Ixodes scapularis</i> |
| B7PE61                  | 26 kDa  | 2                  | 1  | 2  | 1    | Vacuolar assembly/sorting protein DID4   | <i>Ixodes scapularis</i> |
| B7PJ41                  | 78 kDa  | 0                  | 0  | 0  | 2    | Paramyosin, putative (Fragment)          | <i>Ixodes scapularis</i> |
| B7PKP9                  | 36 kDa  | 1                  | 1  | 1  | 0    | Glyceraldehyde-3-phosphate dehydrogenase | <i>Ixodes scapularis</i> |
| B7PSX9                  | 34 kDa  | 0                  | 2  | 0  | 0    | Tropomyosin invertebrate, (Fragment)     | <i>Ixodes scapularis</i> |
| B7PT17                  | 42 kDa  | 0                  | 0  | 0  | 1    | BTB/POZ domain-containing protein        | <i>Ixodes scapularis</i> |
| B7PTL3                  | 40 kDa  | 0                  | 0  | 1  | 2    | Kinesin-related protein HSET             | <i>Ixodes scapularis</i> |
| B7Q1D3                  | 34 kDa  | 0                  | 1  | 0  | 0    | Solute carrier protein, putative         | <i>Ixodes scapularis</i> |
| B7Q2L8                  | 63 kDa  | 0                  | 0  | 1  | 0    | Cdc6 protein, putative                   | <i>Ixodes scapularis</i> |
| B7Q326                  | 48 kDa  | 1                  | 0  | 0  | 0    | Acyl-CoA dehydrogenase, putative         | <i>Ixodes scapularis</i> |
| B7Q381                  | 476 kDa | 2                  | 2  | 0  | 0    | Putative uncharacterized protein         | <i>Ixodes scapularis</i> |
| B7Q5H9                  | 39 kDa  | 1                  | 2  | 1  | 2    | Fructose-bisphosphate aldolase           | <i>Ixodes scapularis</i> |
| B7Q5J9                  | 139 kDa | 2                  | 0  | 0  | 0    | SMC protein, putative                    | <i>Ixodes scapularis</i> |
| B7QBW8                  | 117 kDa | 1                  | 0  | 0  | 2    | Putative uncharacterized protein         | <i>Ixodes scapularis</i> |
| B7QGQ3                  | 12 kDa  | 0                  | 1  | 0  | 0    | Putative uncharacterized protein         | <i>Ixodes scapularis</i> |
| C4YVM4                  | 49 kDa  | 0                  | 1  | 0  | 0    | F pilus assembly protein TraH            | <i>Ixodes scapularis</i> |
| Q4PM63                  | 14 kDa  | 0                  | 2  | 2  | 2    | Histone H2B                              | <i>Ixodes scapularis</i> |

<sup>a</sup> Homologous proteins found by Blast with % of homology, score, e-value and the NCBI reference link are reported too.

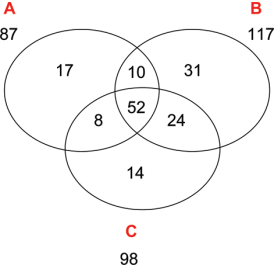


**Figure 3.** Overlapping Venn diagrams of the identification of unique gene products in the three eluates from ProteoMiner. (A) (87 proteins) pH 4.0 eluate, (B) (117 species) pH 7.0 eluate, and (C) (98 proteins) pH 9.5 eluate.

crabs, among which the American species seems to be just a single entity. Possibly the reason for this could be that these arthropods have not been subjected to genome-altering evolutionary pressure seen in other arthropods.

There is even much more in our data that has not been presented here. Although disclosing the identities of 160 protein species has increased the original knowledge of *Limulus* lymph by more than 1 order of magnitude, indeed we have an additional order of magnitude in our data (i.e., another 900 or more proteins) that could not be identified at all, due to lack of homologous sequences in existing databases. We have derived this figure of about 1000 proteins in the lymph by assuming that, for each protein identification, we would need no less than three sequenced peptides, as recommended today.<sup>29</sup> Is this an exaggerated number? Perhaps it is a much too low figure, indeed. Although the lymph in Chelicerata can be considered a primitive form of blood, in human serum/plasma it is believed that several thousands if not hundreds of thousands (considering not just unique gene products, but all possible postsynthetic modifications and the extremely large class of immunoglobulins synthesized by each individual during his life span) could be present.<sup>30</sup> Already in 2004 Anderson et al.<sup>31</sup> merged four different views of the human plasma proteome, based on different methodologies, into a single nonredundant list of 1175 distinct gene products. Soon after, however, with the HUPO plasma proteome project, which enrolled 35 participating laboratories in 13 countries, this number drastically increased to 9504 IPI proteins identified with one or more peptides, which was reduced to a core data set of 3020 proteins identified with two or more peptides.<sup>32</sup> In light of these considerations, a number of 1000 or so unique gene products in the *Limulus* lymph appears quite reasonable and, if anything, underestimated.

Our data also shed light onto novel, important aspects of the CPLL methodology, not so much in terms of spots counted in 2D maps (ca. 200 in control, 890 in CPLL-treated samples, an increment typically found in all previous analyses) and not even in the increment in protein identifications (about twice as many in the CPLL eluates as compared to the control, see Table 1, in line with previous reports of 2 to 5 times as many in all samples so far analyzed). Since the very inception of the CPLL technique, we had always lamented that in the bead eluates, we missed typically 10% (and even higher percentages) of the proteins detected in the control, a drawback of the methodology for which we had not found a remedy. We even hypothesized that baits for the missing proteins might not be present onto the CPLL beads, a hypothesis that would not seem to be tenable, considering the vast bait heterogeneity in

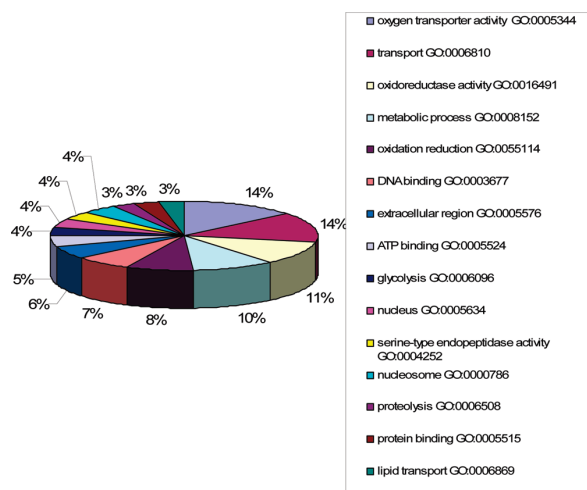


**Figure 4.** Pie chart of the main GO pathways assigned to the proteins identified in the three ProteoMiner eluates of the *Limulus* lymph. The four major pathways are represented by oxygen transport activity (14%), transport processes (14%), oxidoreductase activity (11%), and metabolic processes (10%) (as a representative pie chart, only the GO analysis for the pH 4.0 eluate is shown; all other eluates and control lymph show essentially identical data).

ProteoMiner ligands (millions of diversomers). By surveying Table 1, we can now gather a very important information: barely 1% of the proteins present in the control is missing in the bead eluates, suggesting a major improvement in the performance of the CPLs. There is only one reason for that: today we perform the capture of proteomes not at a single pH value (typically at pH 7.2 in physiological saline) but at 3 different pH values, as suggested by Fasoli et al.<sup>23</sup> As shown in Table 1, it can be appreciated that, in all cases, even if the signal present in the control is not found at all three pH values (although in most cases it is), it is found at least at one of the pH values adopted in the capture. This suggests that the library contains all the possible partners for the proteome under investigation, but that the affinity of the baits is strongly pH-dependent; it might be extremely low at a given pH value and very high at other values of the pH scale. As a consequence of this, it is fair to assume that one of the primary mechanisms for complex formation between a given bait and the partner protein is ion-to-ion interaction, known to be a long-range interaction and among the strongest ones of all noncovalent complexes. Such interactions in fact are the only ones that can be strongly modulated by the prevailing pH in solution. Of course, once the charge on the bait is fully neutralized via ion-pairing with the partner protein, it is quite likely that additional interactions could ensue (e.g., hydrophobic interactions concomitant with hydrogen bonds). It is thus suggested that in all future uses of the CPL technology capture at 3 pH values should be adopted for a proper coverage of any given proteome under investigation.

## 5. Conclusions

Just as the Hubble telescope and the cosmic background explorer satellite (COBE) were set up in orbit to focus on the origin of the universe (the Big Bang and the cosmic background radiation permeating the universe),<sup>33</sup> so we set our ProteoMiner scope back toward the origin of complex life in earth, focusing

on the starting of the Phanerozoic period (500 million years ago). Our data on the *Limulus* lymph indicate the existence of ancestral proteins, which do not seem to be present in today's databases. We could only identify >10% of these ancestral proteins and hope that this report will spur the fantasy and curiosity of geneticists and molecular biologists to map those ancestral genes and bring to life such forgotten ancestral proteins, which might add to the arsenal of novel biochemicals and pharmaceuticals. In addition, our data seem to confirm the relationship among *Limulus* and the other members of the subphylum Chelicerata, such as ticks, scorpions, spiders, and mites, relationships which, according to an extensive study of Shultz,<sup>34</sup> are extremely difficult to establish and should be subjected to extensive revision (although a very recent article seems to confirm the relationship here reported among *Limulus* and Chelicerata).<sup>35</sup>

**Acknowledgment.** P.G.R. is supported in part by Fondazione Cariplo (Milano) and by PRIN 2008 (Rome). We thank Ms. Benjie Lynn Swan of Limuli Laboratories, New Jersey, for supplying horseshoe crabs.

**Supporting Information Available:** Supplementary Table 1 lists the identified proteins in the three eluates and control with access number, molecular mass, number of peptides, peptide sequence, number of spectra, peptide identification probability, and Mascot Ion score. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Crichton, M. *Jurassic Park*; Knopf Publishing Group: New York, 1990; pp 416.
- (2) Woodward, S. R.; Weyand, N. J.; Bunnell, M. DNA sequence from Cretaceous period bone fragments. *Science* **1994**, *261*, 229–232.
- (3) Hedges, S. B.; Schweitzer, M. H. Detecting dinosaur DNA. *Science* **1995**, *268*, 1191–1192.
- (4) Willerslev, E.; Cappellini, E.; Boomsma, W.; Nielsen, R.; Hebsgaard, M. B.; Brand, T. B.; et al. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **2007**, *317*, 111–114.
- (5) Bidle, K. D.; Lee, S.; Marchant, D. R.; Falkowski, P. G. Fossil genes and microbes in the oldest ice on earth. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 13455–13460.
- (6) Salamon, M.; Tuross, N.; Arensburg, B.; Weiner, S. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13783–13788.
- (7) Asara, J. M.; Schweitzer, M. H.; Freimark, L. M.; Phillips, M.; Cantley, L. C. Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* **2007**, *316*, 280–284.
- (8) Buckley, M.; Walker, A.; Ho, S. Y.; Yang, Y.; Smith, C.; Ashton, P.; Oates, J. T.; Cappellini, E.; Koon, H.; Penkman, K.; Elsworth, B.; Ashford, D.; Solazzo, C.; Andrews, P.; Strahler, J.; Shapiro, B.; Ostrom, P.; Gandhi, H.; Miller, W.; Raney, B.; Zylber, M. I.; Gilbert, M. T.; Prigodich, R. V.; Ryan, M.; Rijdsdijk, K. F.; Janoo, A.; Collins, M. J. Comment on protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* **2008**, *319*, 33.
- (9) Pevzner, P. A.; Kim, S.; Ng, J. Comment on protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* **2008**, *321*, 1040b.
- (10) Bern, M.; Brett, S.; Phinney, B. S.; Goldberg, D. Reanalysis of *Tyrannosaurus rex* Mass Spectra. *J. Proteome Res.* **2009**, *8*, 4328–4332.
- (11) Poinar, G. O.; Waggoner, B. M.; Bauer, U. C. Terrestrial Soft-Bodied Protists and Other Microorganisms in Triassic Amber. *Science* **1993**, *259*, 222–224.
- (12) Poinar, G. O., Jr.; Danforth, B. N. A fossil bee from Early Cretaceous Burmese amber. *Science* **2006**, *314*, 614.
- (13) Poinar, G. O., Jr.; Hess, R. Ultrastructure of a 40 Million Year Old Insect Tissue. *Science* **1982**, *215*, 1241–1242.
- (14) Smejkal, G. B.; Poinar, G. O.; Righetti, P. G. Will amber inclusions provide the first glimpse of a Mesozoic proteome? *Expert Rev. Proteomics* **2009**, *6*, 1–4.

- (15) Sekiguchi, K. *Biology of Horseshoe Crabs*; Science House CO.: Tokyo, 1988.
- (16) Sugita, H.; Shishikura, F.; Sugawara, K.; Yonekawa, H.; Tagashira, Y.; Sikiguchi, K. In *Immunological comparison of haemocyanins and their phylogenetic implications*; Sekiguchi, K., Ed.; Biology of Horseshoe Crabs. Science House CO.: Tokyo, 1988; pp 315–334.
- (17) Lamy, J.; Sizaret, P. Y.; Frank, J.; Verschoor, A.; Feldmann, R.; Bonaventura, J. Architecture of *Limulus polyphemus* hemocyanin. *Biochemistry* **1982**, *21*, 6825–6833.
- (18) Guerrier, L.; Righetti, P. G.; Boschetti, E. Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteins by means of hexapeptide ligand library. *Nat. Protoc.* **2008**, *3*, 883–890.
- (19) Castagna, A.; Cecconi, D.; Sennels, L.; Rappsilber, J.; Guerrier, L.; Fortis, F.; Boschetti, E.; Lomas, L.; Righetti, P. G. Exploring the hidden human urinary proteome via ligand library beads. *J. Proteome Res.* **2005**, *4*, 1917–1930.
- (20) Sennels, L.; Salek, M.; Lomas, L.; Boschetti, E.; Righetti, P. G.; Rappsilber, J. Proteomic Analysis of Human Blood Serum Using Peptide Library Beads. *J. Proteome Res.* **2007**, *6*, 4055–4062.
- (21) Roux-Dalvai, F.; Gonzalez de Peredo, A.; Simó, C.; Guerrier, L.; Bouyssié, D.; Zanello, A.; Citterio, A.; Burlet-Schiltz, O.; Boschetti, E.; Righetti, P. G. Monsarrat, B Extensive analysis of the cytoplasmic proteome of human erythrocytes using the Peptide ligand library technology and advanced mass spectrometry. *Mol. Cell. Proteomics* **2008**, *7*, 2254–2269.
- (22) Mouton-Barbosa, E.; Roux-Dalvai, F.; Bouyssié, D.; Berger, F.; Schmidt, E.; Righetti, P. G.; Guerrier, L.; Boschetti, E.; Burlet-Schiltz, O.; Monsarrat, B.; Gonzalez de Peredo, A. In depth exploration of cerebrospinal fluid by combining peptide ligand library treatment and label free protein quantification. *Mol. Cell. Proteomics* **2010**, *9*, 1006–1021.
- (23) Fasoli, E.; Farinazzo, A.; Sun, C. J.; Kravchuk, A. V.; Guerrier, L.; Fortis, F.; Boschetti, E.; Righetti, P. G. Interaction among proteins and peptide libraries in proteome analysis: pH involvement for a larger capture of species. *J. Proteomics* **2010**, *73*, 733–742.
- (24) Candiano, G.; Dimuccio, V.; Bruschi, M.; Santucci, L.; Gusmano, R.; Boschetti, E.; Righetti, P. G.; Ghiggeri, G. M. Combinatorial peptide ligand libraries for urine proteome analysis: investigation of different elution systems. *Electrophoresis* **2009**, *30*, 2405–2411.
- (25) D'Amato, A.; Bachi, A.; Fasoli, E.; Boschetti, E.; Peltre, G.; Sénéchal, H.; Righetti, P. G. In-depth exploration of cows whey proteome via combinatorial peptide ligand libraries. *J. Proteome Res.* **2009**, *8*, 3925–3936.
- (26) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383–5392.
- (27) Binns, D.; Dimmer, E.; Huntley, R.; Barrell, D.; O'Donovan, C.; Apweiler, R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **2009**, *25*, 3045–3046.
- (28) Matafora, V.; D'Amato, A.; Mori, S.; Blasi, F.; Bachi, A. Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition. *Mol. Cell. Proteomics* **2009**, *8*, 2243–55.
- (29) Adamski, M.; Blackwell, T.; Menon, R.; Martens, L.; Hermjakob, H.; Taylor, C. Data management and preliminary data analysis in the pilot phase of the HUPO plasma proteome project. In *Exploring the Human Plasma Proteome*; Omenn, G. S., Ed.; Wiley-VCH: Weinheim, 2006; pp 37–61.
- (30) Anderson, N. L.; Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
- (31) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. The human plasma proteome: a non-redundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **2004**, *3*, 311–326.
- (32) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H. Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. In *Exploring the Human Plasma Proteome*; Omenn, G. S., Ed.; Wiley-VCH: Weinheim, 2006; pp 1–35.
- (33) Chown, M. *Afterglow of Creation*; Faber and Faber: London, 2010.
- (34) Shultz, J. W. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zool. J. Linnean Soc.* **2007**, *150*, 221–265.
- (35) Regier, J. C.; Shultz, J. W.; Zwick, A.; Hussey, A.; Ball, B.; Wetzter, R.; Martin, J. W.; Cunningham, C. W. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **2010**, *463*, 1079–1084.

PR1002033