PLOS ONE

RESEARCH ARTICLE

Transcriptomes reveal expression of hemoglobins throughout insects and other Hexapoda

Hollister W. Herhold*, Steven R. Davis, David A. Grimaldi

Division of Invertebrate Zoology, American Museum of Natural History, New York, New York, United States of America

* hherhold@amnh.org

OPEN ACCESS

Citation: Herhold HW, Davis SR, Grimaldi DA (2020) Transcriptomes reveal expression of hemoglobins throughout insects and other Hexapoda. PLoS ONE 15(6): e0234272. https://doi.org/10.1371/journal.pone.0234272

Editor: Marc Robinson-Rechavi, Universite de Lausanne Faculte de biologie et medecine, SWITZERLAND

Received: February 19, 2020

Accepted: May 21, 2020

Published: June 5, 2020

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors declare that no competing interests exist.

Abstract

Insects have long been thought to largely not require hemoglobins, with some notable exceptions like the red hemolymph of chironomid larvae. The tubular, branching network of tracheae in hexapods is traditionally considered sufficient for their respiration. Where hemoglobins do occur sporadically in plants and animals, they are believed to be either convergent, or because they are ancient in origin and their expression is lost in many clades. Our comprehensive analysis of 845 Hexapod transcriptomes, totaling over 38 Gbases, revealed the expression of hemoglobins in all 32 orders of hexapods, including the 29 recognized orders of insects. Discovery and identification of 1333 putative hemoglobins were achieved with target-gene BLAST searches of the NCBI TSA database, verifying functional residues, secondary- and tertiary-structure predictions, and localization predictions based on machine learning. While the majority of these hemoglobins are intracellular, extracellular ones were recovered in 38 species. Gene trees were constructed via multiple-sequence alignments and phylogenetic analyses. These indicate duplication events within insects and a monophyletic grouping of hemoglobins outside other globin clades, for which we propose the term *insectahemoglobins*. These hemoglobins are phylogenetically adjacent and appear structurally convergent with the clade of chordate myoglobins, cytoglobins, and hemoglobins. Their derivation and co-option from early neuroglobins may explain the widespread nature of hemoglobins in various kingdoms and phyla. These results will guide future work involving genome comparisons to transcriptome results, experimental investigations of gene expression, cell and tissue localization, and gas binding properties, all of which are needed to further illuminate the complex respiratory adaptations in insects.

Introduction

The colonization of land by arthropods in the Paleozoic required profound changes in respiration. These animals transitioned from a system of gills and respiratory proteins, like hemocyanins, to breathing air. Hemoglobins are also well known for their common role in respiration, although they have various other roles too, discussed below. Hemoglobins are globular

proteins of 140–150 aa length, usually comprised of eight 3-over-3 α-helical segments (A-H), with Fe^{++} that binds O_2, CO, NO, and CO_2. The amino acid sequences can vary widely among taxa, but to preserve the oxygen-binding properties there are two highly recognizable regions in hemoglobin, comprising the characteristic "globin fold": Phe and His occur at positions CD1 and F8, respectively, and hydrophobic residues occur in each of the α-helical segments [1]. Hemoglobins occur in at least five kingdoms of life and 12 phyla of animals from protists to vertebrates [1], but despite their apparent antiquity their expression among animals is presently known to be sporadic, and absent in most orders and species of the largest radiation, the insects [1,2].

Hemoglobin expression has been shown to occur in approximately 19 species of insects belonging to 14 families and five orders of insects (Table 1). Additional species have been identified as possessing hemoglobins, using sequence identity (e.g., [3,4]) genome and transcriptome annotations (e.g., [4] and other methods (e.g., [5]). These are not included in Table 1, only hemoglobins that have been functionally, experimentally, or biochemically characterized are listed. Some have been known for many years, particularly in *Chironomus* midge larvae [6–8]; stomach bot flies, *Gasterophilus* [9]; and the backswimmer bugs *Anisops* and *Buenoa* [10,11].

Extensive studies by Burmester and colleagues have revealed the surprising occurrence of hemoglobins in insects that seem to have little or no oxygen limitations, including the well-known *Apis mellifera*, *Drosophila melanogaster*, *Anopheles* mosquitoes, and a dozen species of true bugs, beetles, and moths (Table 1). Most of these characterized hemoglobins are implicated in respiration and associated with the tracheal system, though a few are involved in other physiological and developmental processes (e.g., [1,27,28]). The majority appear to be localized intracellularly, likely cytoplasmic or in the cell membranes of tracheocytes and adipocytes; a few are extracellular, dissolved in the hemolymph and are responsible for its red coloration. Although relatively few insect hemoglobins have been studied in detail, their copy number is variable and appear to show stage- and tissue-specific expression differences [27]. Substantial hemoglobin structural variation also occurs in insects, ranging from monomeric to di-, tetra-, and hexameric quaternary structures. Binding schemes of the heme Fe^{++} in the deoxygenated state varies from penta- to hexacoordinate, the functional implications of which remain unclear.

There are several reasons why, until now, hemoglobins in insects have been thought to occur so sporadically.

1. *Assumptions among most biologists that hemoglobins are entirely respiratory in function*. Although O_2 transport may be the most common function of Hbs, some forms have varied functions, particularly in invertebrates: in the regulation and detoxification of NO [29], acid-base regulation, oxidase-peroxidase activities, and reactions with sulfide and its transport [1]. Other functions relate to O_2 metabolism, such as sensing [30], facilitating O_2 diffusion, and O_2 scavenging [1]. Some hemoglobins, in fact, apparently serve to protect cells against reactive oxygen species (ROS), similar to vertebrate myoglobin. Discontinuous breathing in insects has been shown to be an adaptive response to ROS [31].

2. *Hemocyanins are the primary respiratory proteins in Pancrustacea (Crustacea including insects)* [2]. Hemocyanins are Cu^{+} containing proteins found in some mollusks, Crustacea (including basal insects), and the sister phylum to arthropods, the Onychophora (velvet worms)[2]. They are, so far as known, always freely dissolved in hemolymph, and may be derived from phenoloxidases [2]. They appear to be lost in a clade of Crustacea (Branchiopoda + Copepoda + Thecostraca, where they are replaced by hemoglobin), in the paleopterous insects (Ephemeroptera and Odonata), and in the holometabolans + paraneopterans [2]. Expression of hemocyanin in polyneopteran insects (stoneflies, roaches, mantises,

Table 1. Insects with characterized hemoglobins.

ORDER Genus—Species	Family	Life Stage	Forms	Respiratory?	Location	Fe2+ binding/Structure	Refs.
DIPTERA:							
Chironomus spp.	Chironomidae	larvae	various	Yes	extracellular	Monomers, dimers	[12–16]
Chironomus tentans	Chironomidae	adult	Ctglob1	no?	extracellular		[16–19]
Drosophila melanogaster	Drosophilidae	adult	Dmglob1	probably*	intracellular	Hexa-coordinate/ Monomer	[4,14–16,20,21]
			Dmglob2	no?	intracellular	Monomer	[4,14–16,20,21]
			Dmglob3	no?	intracellular	Monomer	[4,14–16,20,21]
Anopheles gambiae	Culicidae	larva	Agglob1	probably*	intracellular		[15]
			Agglob2	probably*	intracellular		[15]
Aedes aegypti	Culicidae	larva	Aaglob1	probably*	intracellular		[15]
			Aaglob2	probably*	intracellular		[15]
Gasterophilus intestinalis	Oestridae	larva	Giglob1	Yes	intracellular	Pentacoordinate/ Dimer	[4,9,16,22,23]
Glossina morsitans	Glossinidae		Gmglob1	probably*	intracellular		[4]
HYMENOPTERA:							
Apis mellifera	Apidae	all?	Amglob1	probably*	intracellular		[4][16]
HEMIPTERA:							
Acyrthosiphon pisum	Aphididae	all?	Apglob1	probably*	intracellular		[4]
Aphis gossypii	Aphididae	all?	Agglob1	probably*	intracellular		[4]
Anisops assimilis	Notonectidae	all?	Aaglob1	Yes	intracellular	Penta-coordinate/ Monomer, hexamer	[3,16]
Anisops deanei	Notonectidae	all?	Adglob1	Yes	intracellular	Penta-coordinate/ Monomer, hexamer	[3]
			Adglob2	Yes	intracellular	Penta-coordinate/ Monomer, hexamer	[3]
			Adglob3	Yes	intracellular	Penta-coordinate/ Monomer, hexamer	[3]
Buenoa confusa	Notonectidae	all?	Bcglob1	Yes	intracellular	Monomer, dimer	[16,24,25]
Buenoa macrotibialis	Notonectidae	all?	Bmglob1	probably*	intracellular	Monomer, dimer	[3]
Macrocoryxia geoffroyi	Coryxidae	all?	Mgglob1	Yes	intracellular		[16]
Nilaparvata lugens	Delphacidae	all?	Nlglob1	probably*	intracellular		[3]
COLEOPTERA:							
Dascillus cervinus	Dascillidae	all?	Dcglob1	probably*	intracellular		[4,16]
Tribolium castaneum	Tenebrionidae	all?	Tcglob1	probably*	intracellular		[4,16]
LEPIDOPTERA:							
Bombyx mori	Bombycidae	all?	Bmglob1	probably*	intracellular		[4,16,26]
Samia cynthia ricini	Saturniidae	all?	Scglob1	probably*	intracellular		[26]

*Inferred on basis of Hb-producing cells located near or in tracheal cells, or gene identity.

https://doi.org/10.1371/journal.pone.0234272.t001

orthopterans, earwigs, etc.) appears relegated to embryos, so these proteins are thought to be involved either in development or in respiration of the insect embryo among basal lineages of insects.

3. *The adequacy and efficiency of arthropod respiration via tracheae and book lungs.* Every terrestrial group of arthropods has a complex system of invaginations into the body for the direct delivery of air to tissues, principally in the form of tracheae for insects, myriapods (centipedes and millipedes), and some arachnids (e.g., solifugae [32]), or as book lungs in many spiders and all scorpions [33], and pleopod lungs in pill bugs (Oniscoidae) [34]. Because of ambiguous relationships among the 11 orders of arachnids [35], it is difficult to determine exactly how many times arachnids independently evolved tracheae, but based on structure it may have been five times (e.g., [36–38]). The implication of this remarkable, repeated convergence is that tracheae and book lungs are not only necessary for arthropod breathing on land, they are entirely adequate. Indeed, our study was initiated as a result of our work on insect

tracheae. Despite great adaptive variation in the tracheal system among hexapods, we sought to determine if there might be any physiological compensation for differences in respiration.

4. *Insect Hb expression depends upon exceptional life histories where there is strong selection for oxygen-absorption*. Hb imparts the deep red color to certain *Chironomus* larvae that live in hypoxic habitats like the sediments of eutrophic and polluted non-saline water. *Chironomus* has been intensively studied also because the larval hemoglobins are uniquely dissolved in hemolymph rather than within cells, and multiple forms occur in some species (larval *C. tentans*, for example, has up to 40 forms of hemoglobin) [16,39]. Larvae of the horse bot flies, *Gasterophilus*, are embedded into the host stomach wall, obtaining oxygen through air that the horse ingests [9,22]. *Anisops* and some other Corixidae are active predators in ponds, which breathe from bubbles trapped against the body (a plastron) during dives. Their Hb is a source of oxygen later in the dive, so that the plastron is not depleted and the bug can retain buoyancy [11,40]. Although it is not an insect, it merits mention that the only copepod that is known to have hemoglobins, *Benthoxynus spiculifer*, lives in the hypoxic water of deep-sea thermal vents [41].

Here we show through transcriptomics that hemoglobin expression is in fact ubiquitous among insects, including in their hexapod relatives. Based on transcript sequence identity we are able to speculate on the possible function(s) of some, but hemoglobin functions in most of the species will require experimental study. Our results are a rigorous test and confirmation of Burmester's [2](pg. 797) statement that " . . .; apparently many—but probably not all—pancrustacean genomes or transcriptomes harbor an HbL [hemoglobin-like] gene", and the more recent hypothesis [27](pg. 230) that ". . . a glob1-like gene belongs to the standard repertoire of insects . . .". The results have implications for our understanding of globin genes, which are primary models for the evolution of gene duplication and molecular function (e.g., [42]).

Results and discussion

Searching for novel hemoglobin sequences

Previous studies have used bioinformatic methods to investigate the presence of hemoglobins in certain plants, bacteria, and eukaryotes, having focused on individual specimens in order to characterize and determine the functions of single hemoglobins [11,22,43]. The objective of our study was an extensive, comparative approach to test if hemoglobins are present in the genome, but more importantly, if they are expressed across all orders of Hexapoda. Using standard NCBI Transcriptome Shotgun Assembly (TSA) Database online search tools, 845 hexapod transcriptomes, totaling over 38.2 Gbases, were retrieved in June 2018 [44]. In addition to representatives from all 29 recognized insect orders, one transcriptome each from Remipedia and Protura, two from Diplura, and 8 from Collembola were included (See S1 File for identifiers and accession information). Remipedia is a close crustacean relative of hexapods; the latter 3 comprise the Entognatha, which is the living sister group to the Insecta. As 78 taxa had multiple transcriptomes in the database, the final dataset contained a total of 716 species. Species with duplicate transcriptomes were included because gene expression can be dependent on factors such as development stage or tissue type and the aim of the study was to investigate the overall presence of hemoglobins in hexapods.

Potential hemoglobin amino acid sequences were uncovered with a two-tiered BLAST search using insect hemoglobin amino acid sequences from the literature as target genes (S1 Table). Sequences were identified as hemoglobins via multiple sequence alignments to confirm functionally important residues, secondary- and tertiary-structure predictions and alignments to verify the characteristic 3-over-3 α-helix arrangement, and cellular localization predictions to determine probable functionality (see Methods and Materials for overview, and S2 File for

full implementation details on bioinformatic workflow). Fig 1 presents a 60-sequence exemplar subset of identified hemoglobins with representatives from all 32 hexapod orders to illustrate alignment of functionally important residues and secondary structure predictions.

Expressed hemoglobins were located in 681 of 845 transcriptomes, representing 606 species. Although the levels of transcription of a gene do not necessarily dictate how much protein is synthesized[46–48], the broad presence of hemoglobin transcripts throughout hexapods strongly indicates protein expression. While specimen sampling was concentrated in certain orders (Diptera, Hymenoptera, Coleoptera, and Lepidoptera), the analysis uncovered hemoglobins across Hexapoda, with at least one species in every hexapod order possessing transcripts (Table 2). Many species expressed more than one hemoglobin sequence (Fig 2 and S4 File)–which is unsurprising since insects possessing multiple hemoglobin genes are already known [15,21,49]. A total of 1333 unique hemoglobin transcripts were found.

Hexapod transcriptomes per order available from NCBI TSA database, showing number with and without hemoglobins. Note that all orders have at least one hemoglobin-containing transcript. See text regarding explanations for putative lack of hemoglobins in some species. Note that the number of transcriptomes is greater than the number of species for several orders as some species were represented by multiple transcriptomes in the analysis.

Source data—sampling of orders, rejected transcriptomes, contaminated taxa

From the set of 845 transcriptomes, 164 were found to not contain hemoglobins and hemoglobin-like sequences, so we sought to examine this apparent lack. For 36 of these Hb-less transcriptomes, a second transcriptome of the same species contained a hemoglobin transcript, indicating the species indeed contained an expressed hemoglobin. Many of these "double entries" were an updated, larger transcriptome from a smaller, older entry. Sequences located by the search as possible hemoglobins but containing fewer than 80 residues were considered to be unreliable, and this accounted for 16 transcriptomes. The remaining 112 transcriptomes were spread across 12 orders. Each of these 12 orders was represented by another transcriptome that contained hemoglobins; the absence of hemoglobin transcripts in these 112 transcriptomes did not influence the coverage of orders.

The size and quality of the assembled transcriptome was likely a factor in the success of locating hemoglobin. From the remaining 112 transcriptomes, 14 were significantly smaller, by more than an order of magnitude, than the average dataset in both contig and base pair count, and most likely incomplete. An additional 9 transcriptomes were assembled using Roche 454 sequencing, whereas the majority of transcriptomes in the study were assembled using Illumina. It is possible that novel sequences may be missed in 454-sequenced transcriptomes, since this technique has lower throughput.

To address the lack of hemoglobins in the remaining 89 transcriptomes, sample data submitted with each transcriptome's entry in NCBI was reviewed. Hemoglobin is not universally expressed across all tissue types and developmental stages, so targeted sampling used in many studies, such as specific tissues or developmental stages, could affect the presence or absence of hemoglobin transcripts (e.g., [16,27,51]). For example, hemoglobin expression in chironomids is known to be primarily in larval stages, and certain hemoglobins in *Chironomus thummi* are expressed during particular larval instars. Annotations from the NCBI TSA database for each of the remaining transcriptomes were used when supplied, and when no information was submitted, original references were consulted when available.

Forty-seven transcriptomes focused on specific tissue types, such as testes, venom glands, salivary glands, venom itself, and even regurgitant. Although functions of hemoglobins vary

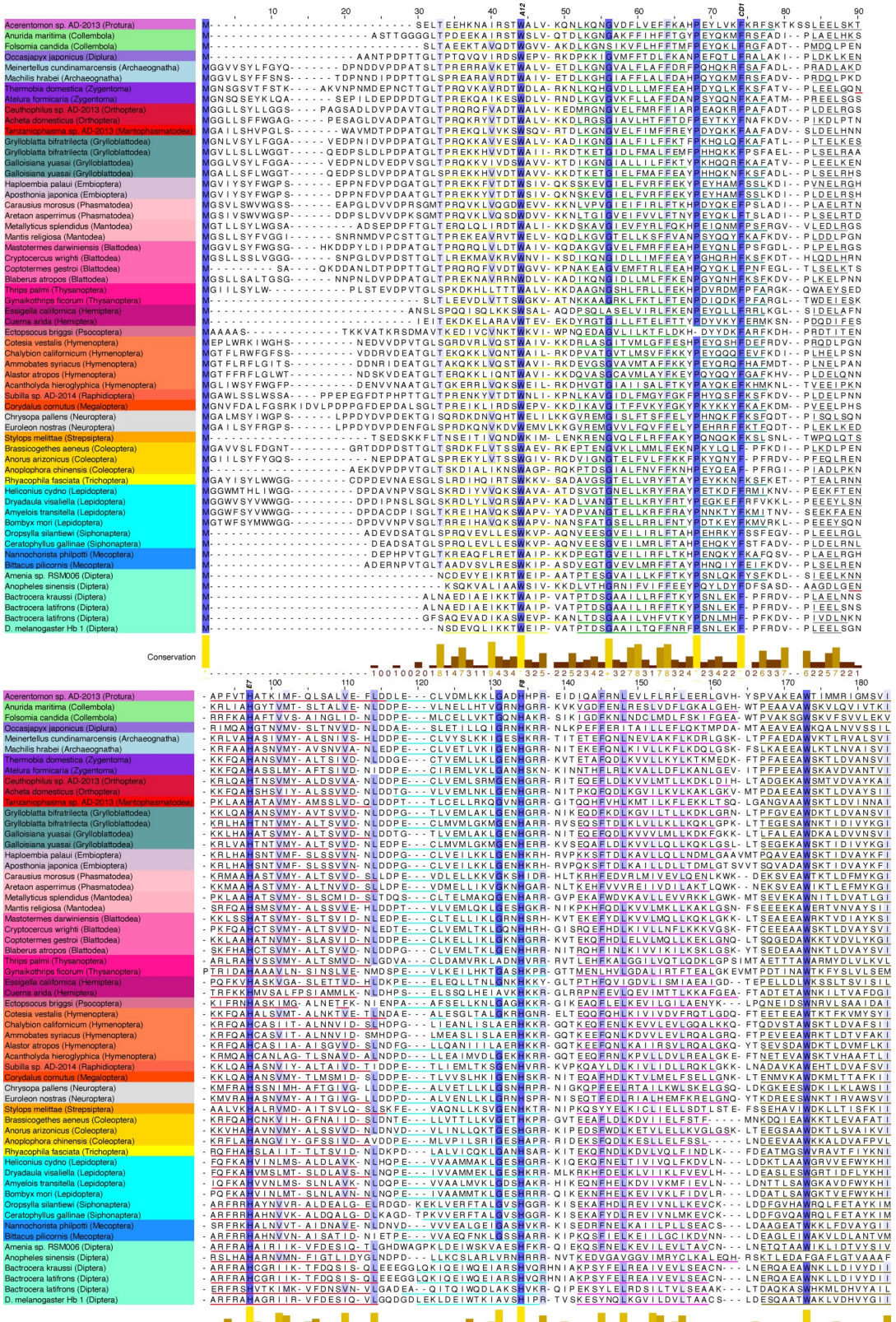
Acerentomon sp. AD-2013 (Protura)
Anurida maritima (Collembola)
Folsomia candida (Collembola)
Occasjapyx japonicus (Diplura)
Meinertellus cundinamarcensis (Archaeognatha)
Machilis hrabei (Archaeognatha)
Thermobia domestica (Zygentoma)
Atelura formicaria (Zygentoma)
Ceuthophilus sp. AD-2013 (Orthoptera)
Acheta domesticus (Orthoptera)
Tanzaniophasma sp. AD-2013 (Mantophasmatodea)
Grylloblatta bifratrilecta (Grylloblattodea)
Grylloblatta bifratrilecta (Grylloblattodea)
Galloisiana yuasai (Grylloblattodea)
Galloisiana yuasai (Grylloblattodea)
Haploembia palaui (Embioptera)
Aposthonia japonica (Embioptera)
Carausius morosus (Phasmatodea)
Aretaon asperrimus (Phasmatodea)
Metallyticus splendidus (Mantodea)
Mantis religiosa (Mantodea)
Mastotermes darwiniensis (Blattodea)
Cryptocercus wrighti (Blattodea)
Coptotermes gestroi (Blattodea)
Blaberus atropos (Blattodea)
Thrips palmi (Thysanoptera)
Gynaikothrips ficorum (Thysanoptera)
Essigella californica (Hemiptera)
Cuerna arida (Hemiptera)
Ectopsocus briggsi (Psocoptera)
Cotesia vestalis (Hymenoptera)
Chalybion californicum (Hymenoptera)
Ammobates syriacus (Hymenoptera)
Alastor atropos (Hymenoptera)
Acanthoylda hieroglyphica (Hymenoptera)
Subilla sp. AD-2014 (Raphidioptera)
Corydalus cornutus (Megaloptera)
Chrysopa pallens (Neuroptera)
Euroleon nostras (Neuroptera)
Stylops melittae (Strepsiptera)
Brassicogethes aeneus (Coleoptera)
Anorus arizonicus (Coleoptera)
Anoplophora chinensis (Coleoptera)
Rhyacophila fasciata (Trichoptera)
Heliconius cydno (Lepidoptera)
Dryadaula visaliella (Lepidoptera)
Amyelois transitella (Lepidoptera)
Bombyx mori (Lepidoptera)
Oropsylla silantiewi (Siphonaptera)
Ceratophyllus gallinae (Siphonaptera)
Nannochorista philpotti (Mecoptera)
Bittacus pilicornis (Mecoptera)
Amenia sp. RSM006 (Diptera)
Anopheles sinensis (Diptera)
Bactrocera kraussi (Diptera)
Bactrocera latifrons (Diptera)
Bactrocera latifrons (Diptera)
D. melanogaster Hb 1 (Diptera)
Conservation

Fig 1. Hemoglobin exemplar MSA. Multiple sequence alignment of hemoglobin amino acid sequences selected from 60 exemplar species across Hexapoda, truncated at 180 residues (end of helix H) to conserve space. Background colors correspond to specimen orders (Diptera, etc) for all subsequent figures. Secondary structure predictions are annotated using colored underlines to indicate α-helices corresponding to helices detailed in S3 File. Helix colors are as follows: A = yellow, B = light green, C = dark green, D = gray, E = red, F = teal, G = magenta, and H = brown. Locations of functionally important residues (in particular, tryptophan at A12, phenylalanine at CD1, and histidine at E7 and F8) are noted as blue vertical bars. Conservation annotations computed by JalView, measuring the number of physio-chemical properties conserved [45].

https://doi.org/10.1371/journal.pone.0234272.g001

Table 2. NCBI hexapod transcriptomes.

Order	Transcriptomes	Transcriptomes with Hbs	Transcriptomes with no Hbs	Species	Species with Hbs	Species with NO Hbs
Protura	1	1	0	1	1	0
Collembola	8	8	0	7	7	0
Diplura	2	1	1	2	1	1
Archaeognatha	2	2	0	2	2	0
Zygentoma	3	3	0	3	3	0
Odonata	7	7	0	7	7	0
Ephemeroptera	5	5	0	5	5	0
Zoraptera	2	2	0	2	2	0
Dermaptera	3	3	0	2	2	0
Plecoptera	4	3	1	4	3	1
Orthoptera	20	16	4	16	13	3
Mantophasmatodea	1	1	0	1	1	0
Grylloblattodea	2	2	0	2	2	0
Embioptera	2	2	0	2	2	0
Phasmatodea	21	21	0	18	18	0
Mantodea	3	2	1	3	2	1
Blattodea	12	10	2	9	8	1
Thysanoptera	5	3	2	5	3	2
Hemiptera	78	57	21	55	43	12
Psocoptera	2	2	0	2	2	0
Phthiraptera	1	1	0	1	1	0
Hymenoptera	284	257	27	266	248	18
Raphidioptera	3	2	1	3	2	1
Megaloptera	3	2	1	3	2	1
Neuroptera	7	7	0	7	7	0
Strepsiptera	2	1	1	2	1	1
Coleoptera	71	54	17	56	44	12
Trichoptera	6	6	0	6	6	0
Lepidoptera	131	87	44	90	64	26
Siphonaptera	4	3	1	4	3	1
Mecoptera	4	4	0	4	4	0
Diptera	146	106	40	126	97	29
Totals:	845	681	164	716	606	110

https://doi.org/10.1371/journal.pone.0234272.t002

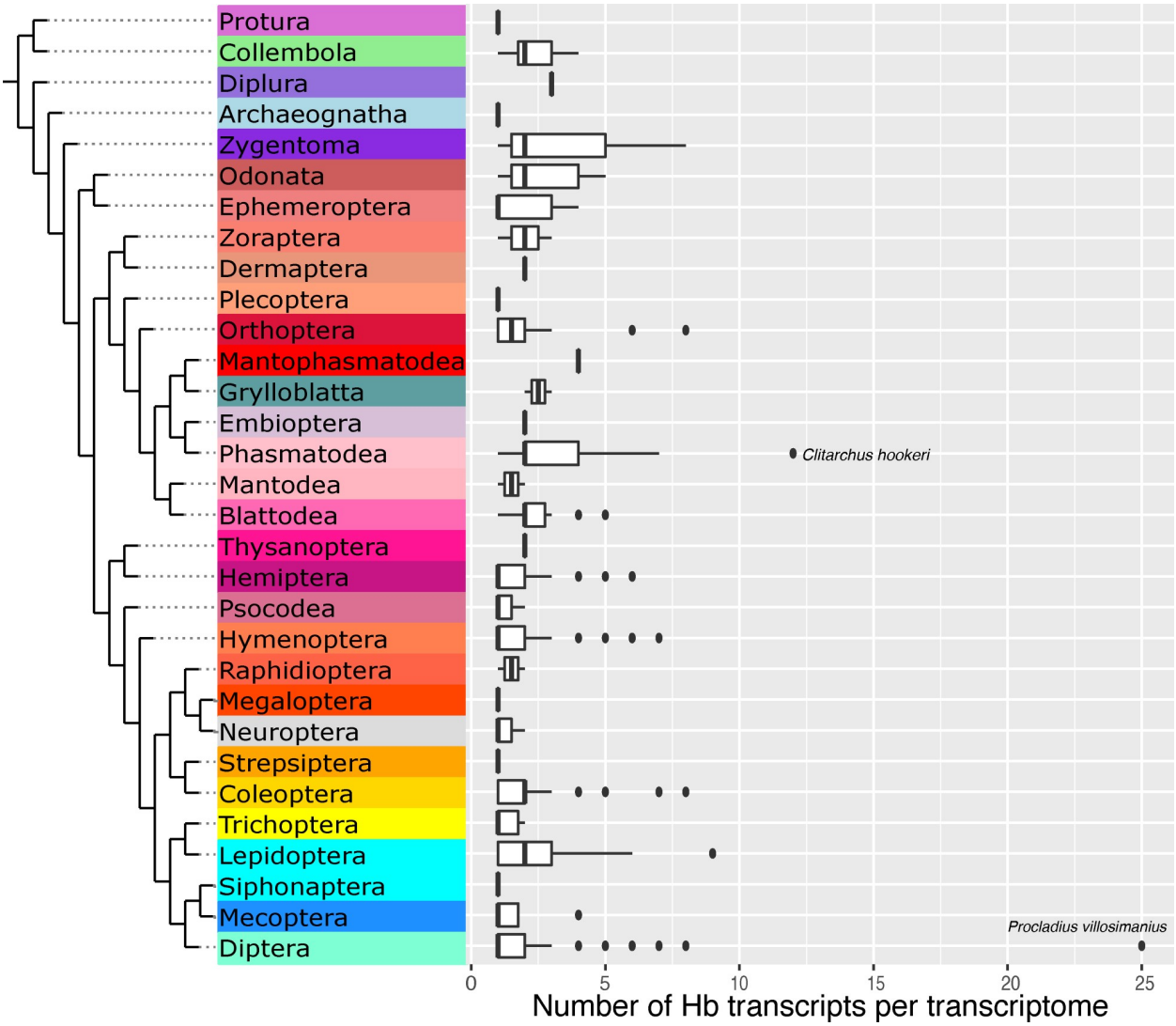

Fig 2. Hexapod phylogeny with transcripts per transcriptome per order. Box and whisker plot of the number of hemoglobins in each transcriptome, sorted by order. Box represents interquartile range with bar at the median value, whiskers indicate 1.5 x interquartile range, dots are outliers. Note, for example, *Procladius villosimanus*, a chironomid midge known to have multiple hemoglobin genes. Phylogeny after Misof et al. [50].

https://doi.org/10.1371/journal.pone.0234272.g002

widely, in insects it is likely that sampling specific tissues could affect the search for hemoglobin transcripts. Additionally, 13 transcriptomes were taken from specific developmental stages. Since our study is a broad comparison for hemoglobins, transcriptomes assembled from a wide array of developmental stages—egg, larva, pupa, and adult—were considered to be more reliable datasets.

We were unable to assess the apparent lack of hemoglobin in 29 transcriptomes by examining metadata from NCBI and reviewing supplied references. These transcriptomes were spread across the following nine orders: Diplura, Orthoptera, Hemiptera, Hymenoptera, Raphidioptera, Coleoptera, Strepsiptera, Lepidoptera, and Diptera. (Table 3). We could find no apparent ecological, phylogenetic, or other commonalities to taxa without expressed hemoglobin transcripts. Further studies should test these apparent absences, ideally using both experimental and bioinformatic approaches.

Table 3. Taxa with no Hb transcripts located.

Taxon	Prefix	Order	Family
Nicrophorus orbicollis	GGAA01	Coleoptera	Silphidae
Nicrophorus vespilloides	GDKQ01	Coleoptera	Silphidae
Campodea augens	GAYN02	Diplura	Campodeidae
Bombylius major	GATI02	Diptera	Bombyliidae
Verticia nigra	GGHV01	Diptera	Calliphoridae
Chaoborus flavidulus	GGBK01	Diptera	Chaoboridae
Toxorhynchites sp. Toxo	GGBL01	Diptera	Culicidae
Tripteroides aranoides	GGBM01	Diptera	Culicidae
Liogma simplicicornis	GEMK01	Diptera	Cylindrotomidae
Bixinia sp. RSM007	GGGY01	Diptera	Rhinophoridae
Tipula maxima	GACZ01	Diptera	Tipulidae
Sitodiplosis mosellana	GAKJ01	Diptera	Cecidomyiidae
Acanthosoma haemorrhoidale	GAUV02	Hemiptera	Acanthosomatidae
Diaphorina citri	GACJ01	Hemiptera	Liviidae
Murgantia histrionica	GECQ01	Hemiptera	Pentatomidae
Courtella sp. AD-2014	GBTH01	Hymenoptera	Agaonidae
Elisabethiella stueckenbergi	GBTW01	Hymenoptera	Agaonidae
Aphidius ervi	GFLW01	Hymenoptera	Braconidae
Diaeretus essigellae	GBWM01	Hymenoptera	Braconidae
Orussus abietinus	GAUJ02	Hymenoptera	Orussidae
Orussus unicolor	GBTS01	Hymenoptera	Orussidae
Megaphragma amalphitanum	GFME01	Hymenoptera	Trichogrammatidae
Acanthopteroctetes unifascia	GENP01	Lepidoptera	Acanthopteroctetidae
Agathiphaga queenslandensis	GENX01	Lepidoptera	Agathiphagidae
Pseudopostega quadristrigella	GEOT01	Lepidoptera	Opostegidae
Plutella xylostella	GFRV01	Lepidoptera	Plutellidae
Locusta migratoria	GEZB01	Orthoptera	Acrididae
Raphidia ariadne	GACX01	Raphidioptera	Raphidiidae
Mengenilla moldrzyki	GACY01	Strepsiptera	Mengenillidae

https://doi.org/10.1371/journal.pone.0234272.t003

The NCBI data was generally considered to be accurate and reliable, but an interesting example of contamination was uncovered. During the phylogenetic analysis (discussed below), a single sequence from the lacewing *Pseudomallada prasinus* (Neuroptera: Chrysopidae) was found to fall inside a clade of parasitoid wasps from the Braconidae and Ichneumonidae (See S1 Fig for the full amino acid cladogram). It is highly unlikely that the lacewing possesses hemoglobins similar in sequence to wasps; in fact, this transcriptome contains a second hemoglobin sequence that clusters with other Chrysopidae sequences. The most likely explanation is that the selected *P. prasinus* specimen harbored an endoparasitoid wasp larva (wasps of the families Eupelmidae and Perilampidae, for example, are known to be parasitoids of Chrysopidae). This brings up a few interesting points, including the possible transcriptome contamination of other parasitized specimens, and the apparent presence of hemoglobin in the wasp egg or larva.

Insectahemoglobins (IHbs)—establishing homology and defining orthologs

As discussed, hemoglobin is present in a wide spectrum of life. The predominant functions of hemoglobins are enzymatic. Oxygen transport is a specialized development that accompanied

the evolution of metazoans [19]. Hemoglobin sequences have diverged significantly to accommodate various roles [1], with many other functions likely to be uncovered (e.g., [4], [27]). Indeed, compared to vertebrate hemoglobins, those of invertebrates show much greater sequence and structural variation [1], perhaps just a reflection of the enormous diversity of the latter in terms of species numbers, body plans, physiology, and life histories.

The three hemoglobins of *Drosophila melanogaster* have been characterized in great detail fairly recently (e.g., [14,20,21,27]). The function of its globin-1 gene is likely to be for respiration and its glob-2 and 3 genes for male reproduction and spermatogenesis, so these have served as templates for finding orthologous genes in other taxa. According to primary sequence, including the presence of conserved sites (Figs 1 and S2 and S4 and S5 Files) and secondary structure, the majority of hexapod hemoglobins appear to be homologs of the *D. melanogaster* globin-1 gene (and other insect glob1 genes that have been described, sensu Table 1) or similar to the hemoglobin-like globins of Blank and Burmester [52] and Burmester et al. [27]. Although the globin-1 gene in brachyceran Diptera (including *D. melanogaster*) has an abbreviated N-terminal region before the first alpha-helix, this appears to be a derived condition, while most other hexapod globin-1 copies possess a longer N-terminal region. From our analyses of transcriptomes, it is unclear if the globin-2 and globin-3 copies in *D. melanogaster* are present in other insects as well. Many of the clades indicative of gene copies could represent independent acquisitions of a globin-2/3 function in various orders, and as has been shown in avian hemoglobins [42], similar biochemical functions can be acquired in independent ways. Particularly if androglobins [53] have been lost in many insect groups, as has been hypothesized for *D. melanogaster* [27], then globin-2/3 functions may have been independently acquired in several insect groups in compensation. As Gleixner et al. [51] demonstrated, the *D. melanogaster* globin-2/3 genes are quite divergent and have evolved at faster rates following duplication. Congruent with their findings, a small clade of dipterans (including *D. melanogaster*) apparently possessing globins-2/3 appear distant from the larger globin-1 clade in the cladogram (Figs 3 and S1) and may be a result of long-branch attraction. Interestingly, this hemoglobin 2/3 clade, in addition to *D. melanogaster*, contains only some mosquitoes, taxa that are not at all closely related within the Diptera. Within the monophyletic clade of hexapod hemoglobins, termed here ***insectahemoglobins (IHBs)***, the appearance of orders belonging to various hexapod clades (e.g., Entognatha, Palaeoptera, Polyneoptera) into two distinct parts of the tree indicate that transcripts are derived from at least two or more insectahemoglobin genes (as is the case for several orders that appear several times in separate parts of the tree, e.g., Zygentoma, Odonata, Blattodea, Phasmatodea, Hemiptera, some Coleoptera and Diptera) (S2 Fig). Analogous to the appearance of glob2-3 in Diptera, gene duplication events appear rather sporadic and could be specific to the order, family, or perhaps lower taxonomic levels. Distinct globin clades correspond to Blank and Burmester's [52] globin X and globin X-like groupings (Figs 3 and S1). As the cladogram was estimated using transcripts from several gene copies, discrepancies with established hexapod phylogeny were expected (S3 Fig). Judging from localization predictions, certain phylogenetic discrepancies (e.g., Hemiptera adjacent to Diptera) likely also result from homoplasy due to convergent functions. However, basic relationships were conserved surprisingly well among groups, particularly within orders that likely contain only insectahemoglobin transcripts and their splice variants (S4–S6 Figs).

Although transcriptomes were not available for all of the species with previously described hemoglobins (Table 1), we recovered nearly all hemoglobin sequences for those which were, including a sequence likely from a novel gene copy of the *Bombyx mori* globin-1. The few cases of missing sequences from described hemoglobins appears to be attributable to the reasons mentioned above for missing hemoglobin sequences in various transcriptomes.

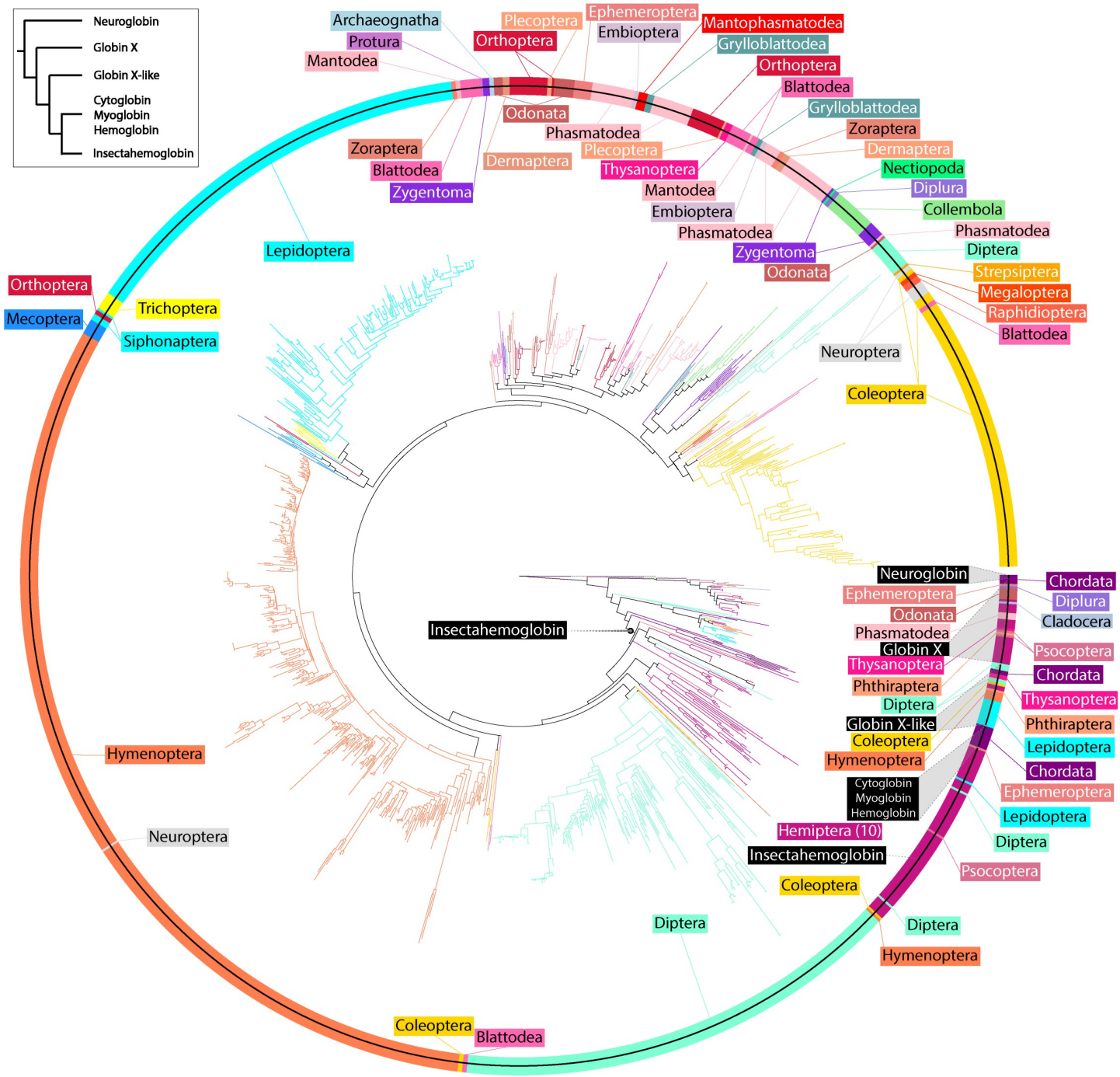

Fig 3. Circular hemoglobin gene tree of all 1382 coding sequences produced with maximum likelihood inference, denoting relationships of different globin types across chordates and the 32 hexapod orders. Colors are order specific, the same as used in Fig 2. Relationships among orders roughly correspond with the current understanding of hexapod relationships, but not directly due to the presence of transcripts from globins X and X-like and those resulting from gene duplications of hemoglobins (i.e., corresponding to *D. melanogaster* globins 1, 2, and 3). A simplified depiction of the globin relationships in the circular tree is presented in the top left inset.

https://doi.org/10.1371/journal.pone.0234272.g003

We further classified some hexapod globins by incorporating into the analyses globins annotated in NCBI as hemoglobins, cytoglobins, neuroglobins, and even one supposed

myoglobin (Figs 4 and S1 Fig). While the sequence terminology in such cases probably is derived from auto-annotated pipelines and inappropriate, the majority of the cytoglobins and neuroglobins actually fall within the globin X and X-like clades. Additionally, secondary structure was estimated for all globin sequences for visual comparison (Fig 1 and S3 Fig, S3 and S4 Files). Many groupings within and among insect orders corresponded to similarities in secondary structure. Structural differences within these groupings is probably due to the presence of additional gene copies or splice variants in various taxa. For instance, as mentioned the

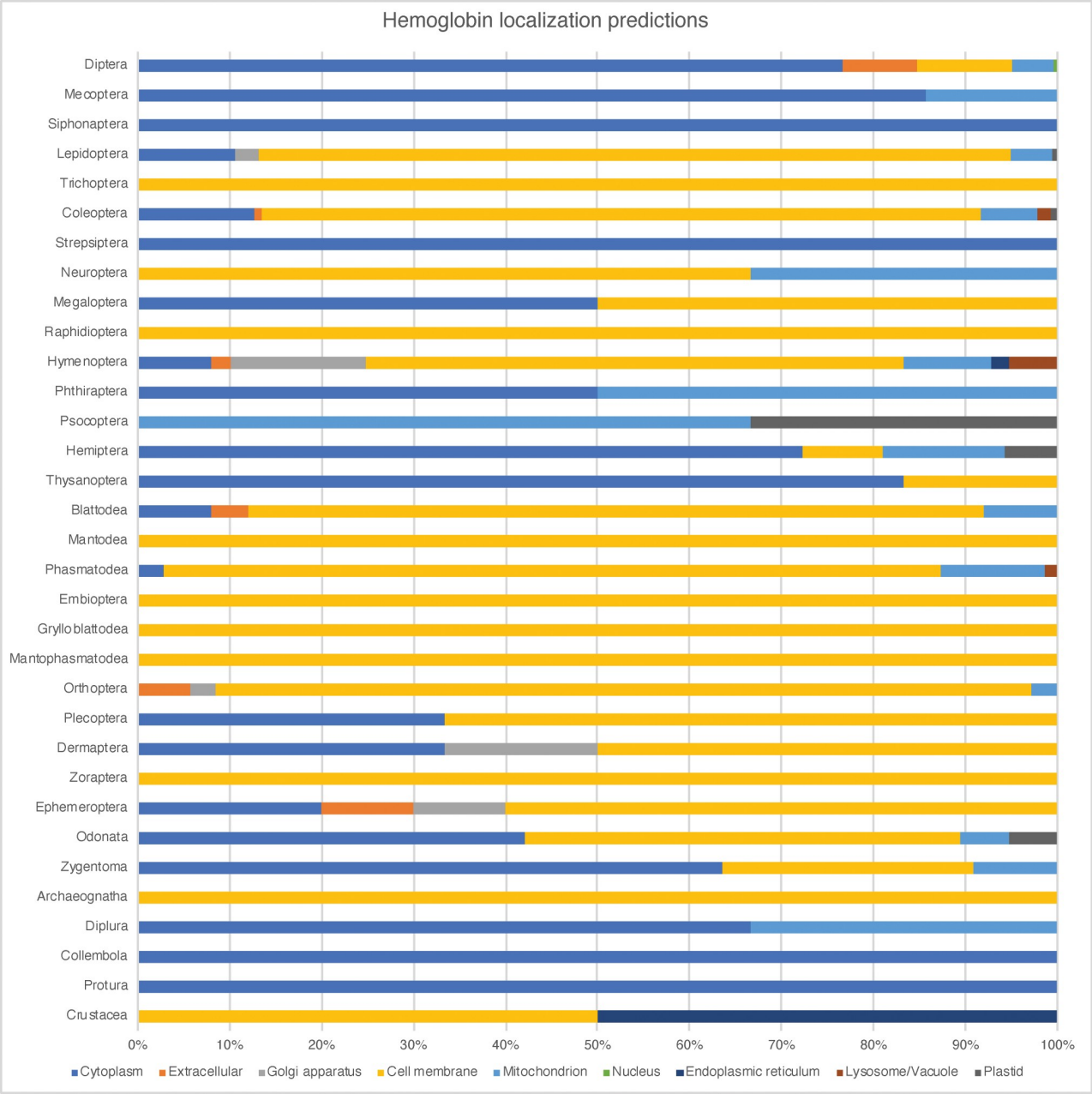

Fig 4. Localization predictions. Stacked bar plot of DeepLoc results, displaying globin localization predictions by order (S6 File for full results on all taxa). The majority of hexapod globins are intracellular and localized to the cell membrane (yellow), with many also found in the cytoplasm (dark blue) followed by the mitochondria (light blue).

https://doi.org/10.1371/journal.pone.0234272.g004

brachyceran Diptera possess a fairly abbreviated IHb with a short N-terminal domain (Fig 1 and S3 Fig). Lepidoptera possess a rather uniform IHb with a long N-terminal region, sometimes containing a short alpha-helix, possibly indicating different splice variants or stemming from different gene copies. Polyneopterans (earwigs, grasshoppers, roaches [including termites], mantises, and others) show some variation in their IHb but mostly have a long N-terminal region with a short alpha-helix. Coleoptera mostly show a long N-terminal domain containing loops, and Hymenoptera have IHb copies also with a long N-terminus containing either only loops or a short alpha-helix. The IHbs in Hemiptera are perhaps most variable in secondary structure in possessing transcripts with N-terminal domains of various lengths and alpha-helices of various lengths.

These patterns in IHb secondary structure among hexapod orders roughly correspond to those seen in our cell-localization predictions, in which the majority of dipteran and hemipteran IHbs are localized in the cytoplasm, while the majority of those in Hymenoptera, Lepidoptera, and Coleoptera are localized to cell membranes (Fig 4 and S6 File). Fewer IHbs are restricted to the Golgi apparatus, mitochondria, or which are extracellular.

To characterize protein structural differences, protein structure predictions were generated for all hemoglobins using I-TASSER and protein alignments estimated for each hemoglobin against all others using TM-align (S7 Fig and S7 and S8 Files). Generally, in congruence with the phylogenetic results of this study (including current data on hexapod phylogeny) and as noted above, IHb proteins were most similar (according to TM-score) within Hymenoptera and Diptera, closely followed by Coleoptera and Lepidoptera. Amphiesmenopteran (Lepidoptera and Trichoptera) scores were highest in comparison to the Antliophora (Diptera, Siphonaptera, and Mecoptera), the two groups comprising the Panorpida, and polyneopteran TM-scores were highest among all groups outside of Holometabola excepting Hymenoptera.

Evolution of hexapod hemoglobins

Recent studies have hypothesized that hemoglobins may be present in most or all hexapods. Here we demonstrate this is the case and it involves a complex evolutionary history of gene duplication and several globin types (globin X, globin X-like, and insectahemoglobin). The relationships presented here among these globin types are congruent with previous studies (e.g., [52], [54–56]). As discussed by Blank and Burmester [52], globin X (and likely globin X-like) may not be involved in respiration. Our results show a scattered sampling of insect taxa in these two globin clades (Fig 3 and S1 Fig), containing transcripts that are membrane bound, cytoplasmic, and localized to mitochondria, probably indicative of various functions. Adjacent to the chordate myoglobins, cytoglobins, and hemoglobins, the remaining hexapod clades fall within the monophyletic insectahemoglobins, the majority derived from glob1-like gene copies. The phylogenetic location of IHbs suggests a structural convergence to the three chordate globins above. These results are consistent with previous studies that postulate independent acquisitions of hemoglobin function from early neuroglobin precursors through gene duplication and co-option [52]. This phenomenon may explain the widespread occurrence of hemoglobins in various kingdoms and phyla. The presence of key amino acids in IHbs indicates an ability to bind O_2, but whether their sole function is respiratory is uncertain. Because of the divergence of the *D. melanogaster* glob2/3 genes and the current uncertainty in taxon distribution of androglobins in hexapods, it is unclear which other hexapods (if any) have independently evolved gene copies with glob2/3-like functions among the IHbs identified in this study. It therefore is possible that some fraction of the IHb genes are copies with glob2/3-like functions. Indeed, the current demonstration of hemoglobin expression in all hexapods leads to questions regarding relations to life histories (e.g., immatures versus adults, aquatic versus

terrestrial, high versus low altitude), development, or various metabolically expensive processes such as flight, running, and swimming. It is difficult to determine any possible relationships of hemoglobin presence/transcript number to tracheal system architecture, which was the impetus for this study. It is possible that hemoglobins are common to nearly all life to support a respiratory role. Perhaps in insects, IHbs have a role in sequestration and intracellular transport and metabolism of respiratory gases, regardless of the tracheal architecture. The novel hemoglobin found in Remipedia in this study suggests IHbs may also be more prevalent in Arthropoda than previously recognized. Given the ubiquity of IHbs throughout hexapods, and their multiple forms (some proven to be important in respiration), the assumption that hexapod respiration is an entirely mechanical process involving just gaseous diffusion through tracheae probably needs to be abandoned. Future studies will need to address Hb functions in various taxa in attempting to define paralogs within the IHbs and to make broad-scale comparisons in the context of hexapod evolution. These analyses and results demonstrate the utility of large-scale genomic and transcriptomic computational analyses in guiding future experimental studies that will, in this case, probe the tissues and intracellular locations of hemoglobins and their functions.

Materials and methods

Custom tools were developed in Python [57] and R [58] for data gathering, searches, analyses, and visualizations. Python versions 3.6.1 and 3.7.2 were used for scripting, in conjunction with BioPython versions 1.69 and 1.72 [59]. R version 3.5.1 was used within R Studio 1.1.453 [60]. The Supplementary Information contains implementation details for all scripts used in the study and a more comprehensive discussion of the workflow.

Searching for hemoglobins

Candidate transcriptomes were obtained from the NCBI Transcriptome Shotgun Assembly (TSA) database [44]. Using online search tools, 845 taxa were identified, containing specimens from 29 orders of Insecta and 3 orders of Entognatha (S1 File). All transcriptomes were downloaded to local storage for analyses performed on-site at the American Museum of Natural History. BLAST searches and subsequent data analyses were run locally on a 16-node high-performance computing cluster using 256 cores. Although jobs were conducted in parallel as much as possible, more than 4 months of wallclock time was used in the analysis.

Using tblastx [61], candidate hemoglobin sequences were identified by comparing eight established arthropod (mostly insect) hemoglobins (S1 Table) against the transcriptomes of all 845 taxa. Although more computationally expensive than other BLAST searches, as all reading frames are checked, tblastx is effective at locating more conserved regions, and can be useful in locating novel genes in insect transcriptomes. Sequences with an Expect value greater than 0.01 were rejected.

All matched sequences were further verified and filtered using blastp to compare against the complete non-redundant (nr) NCBI protein database [62]. Sequences with high Expect values (greater than 0.01) and not matching a keyword search of 'globin' were discarded. Using the verified matches, which were occasionally "fragments", full coding sequences were extracted from the local transcriptome databases using custom-developed tools. Both nucleotide and amino acid datasets were constructed.

Homology assessment

To properly identify homologies with established hemoglobins, 48 known globin sequences (see S2 Table) sampled from across Hexapoda and Chordata were included in a multiple

sequence alignment. In addition to hemoglobins, previously annotated neuroglobins and cytoglobins were obtained from NCBI and included to better annotate the phylogenetic analysis and determine similarities to non-hemoglobin globins. To obtain a proper alignment, 92 short sequences consisting of fewer than 80 residues were removed. In addition, 29 sequences were hand-edited to truncate extraneous residues upstream of the start codon that were included due to incorrect automatic ORF determinations, and 10 additional sequences were removed as they were unable to be properly aligned. Removal of these sequences did not result in the reduction of representation across hexapod orders. The final dataset contained 1333 candidate hemoglobin sequences and the 48 globins from S2 Table, for a total of 1381 sequences. Alignments and visualizations of the 60-sequence subset (S5 File) chosen as "exemplars" used AliView [63], MAFFT [64], and the ETE toolkit [65].

The amino acid dataset was aligned using MAFFT 7.310 with the Needleman-Wunsch algorithm and 1000 cycles of iterative refinement. RevTrans 1.4, a codon-based aligner, was used for the nucleotide dataset [66]. The resulting alignments were 865 positions for amino acids (S4 File) and 2463 for nucleotides.

Phylogenetic analyses

Hemoglobin gene trees were constructed using RAxML 8.2.11 [67]. Amino acid coding sequence relationships were determined using the LG substitution model with empirical base frequencies, optimization of substitution rates with a GAMMA model of rate heterogeneity, and 1000 bootstrap replicates with a rapid bootstrap analysis. For the nucleotide analysis, GTR with optimization of substitution rates and a GAMMA model of rate heterogeneity was used (alpha parameter was estimated), along with an estimate of proportion of invariable sites.

Structure prediction and alignment

Secondary structure prediction on all 1381 sequences used a local installation of PSIPRED 4.0 [68] and the Uniref90 [69] database. Additionally, full three-dimensional structure models were computed using both a local installation of I-TASSER 5.1 and the on-line I-TASSER server [70,71]. Tertiary structure alignments were performed for every predicted model against every other model using TM-align [72].

Localization predictions

Cellular localization prediction was achieved using the online DeepLoc 1.0 server.[73] Results were processed using custom Python scripts and visualizations created with Excel. Predictions were further verified using the NCBI Conserved Domains Database online search.[74]

Other tools

Tree visualizations were composed using the ETE toolkit.[65] Conservation of functionally important residues was calculated using Jalview 2.10.5 [45]. AliView 1.24 was used for viewing and editing alignments and sequences. Mesquite version 3.51 was used for viewing alignments, formatting files for use by RAxML and other programs, and previewing trees [75]. Tanglegrams were rendered using the ETE toolkit and Abobe Illustrator CC 2018 (Version 22.1). Heat maps from tertiary alignments were produced using Python via JupyterLab and the libraries Pandas, Seaborn, MatplotLib, and Numpy from the SciPy toolkit [76,77].

Supporting information

S1 Fig. Amino acid globin tree with chordate globins. Cladogram of all globin transcripts, including 18 chordate globins, from amino acid analysis in RAxML. Bootstrap values labeled at nodes. Colored tags near various taxa in tree indicate locations of previously known hemoglobin sequences included to assist in homology assessment (S2 Table)['globin', red]. Hemoglobin types labeled as belonging to globin X and X-like (according to Blank and Burmester [52]) or insectahemoglobin lineages (i.e., corresponding to D. melanogaster globins 1, 2, and 3).
(PDF)

S2 Fig. Insectahemoglobin gene tree, rooted on Remipedia.
(PDF)

S3 Fig. Globin gene tree and phylogeny of insects. Tanglegram comparing the full globin tree (left), resulting from the nucleotide analysis of all hexapod hemoglobin transcripts in this study (summarized by order), and the phylogeny of Hexapoda (right) (redrawn from Misof et al. [50]). Note that phylogeny includes insectahemoglobins and X and X-like globins. Relative topological incongruence between gene (globin) and taxon trees is indicated by number of overlapping lines connecting clades.
(PDF)

S4 Fig. Diptera hemoglobins and corresponding Diptera family phylogeny. Tanglegram comparing the tree of Diptera hemoglobins (left) and the phylogeny of Diptera (right) (redrawn from Wiegmann et al. [78]). IHb relationships are derived from the full taxon amino acid analysis.
(PDF)

S5 Fig. Hymenoptera hemoglobins and corresponding Hymenoptera family phylogeny. Tanglegram comparing the Hymenoptera insectahemoglobin tree (left) and the phylogeny of Hymenoptera (right) (redrawn from Peters et al. [79]). IHb relationships are derived from the full taxon amino acid analysis.
(PDF)

S6 Fig. Lepidoptera hemoglobins and corresponding lepidoptera family phylogeny. Tanglegram comparing the Lepidoptera insectahemoglobin tree and the phylogeny of Lepidoptera (redrawn from Mitter et al. [80]). IHb relationships are derived from the full taxon amino acid analysis.
(PDF)

S7 Fig. Globin structural similarity. Heatmap produced from protein prediction structure alignments. Structure predictions from I-TASSER were used to compare every globin against every other globin using TM-align. Comparisons are not transitive–above the diagonal, the row element taxon is the template protein, below the diagonal, the column element taxon is the template protein. Values along the diagonal are 1.0. Nearly all TM-align values were greater than 0.5, the threshold for structural similarity [72]. Taxa are listed in phylogenetic order after Misof et al. [50]. Shade of green indicates similarity. White areas indicate missing structure predictions.
(PNG)

S1 Table. Hemoglobin "target" genes used for searching.
(DOCX)

S2 Table. Established 'globin' genes used for alignments and phylogenetic analysis.
(DOCX)

S1 File. TSA Hexapoda transcriptome data table.
(CSV)

S2 File. Supplementary methods and materials. Workflow and scripting implementation details.
(DOCX)

S3 File. Globin transcript multiple sequence alignment (amino acids).
(FASTA)

S4 File. Hemoglobin counts per taxon.
(CSV)

S5 File. Multiple sequence alignment of 60 exemplar hemoglobin amino acid sequences selected from across Hexapoda.
(FASTA)

S6 File. DeepLoc results for all globin sequences.
(XLSX)

S7 File. TM-align output of protein alignments between all hemoglobins.
(XLSX)

S8 File. TM-align output of protein alignments between hemoglobins of exemplar subset.
(XLSX)

Acknowledgments

Special thanks to Apurva Narechania and Sajesh Singh for their assistance with running our analyses on the AMNH high-performance computing clusters. We also thank one anonymous reviewer and Dr. Pierre Kerner, whose suggestions helped improve this manuscript.

Author Contributions

Conceptualization: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Data curation: Hollister W. Herhold, Steven R. Davis.

Formal analysis: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Investigation: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Methodology: Hollister W. Herhold, Steven R. Davis.

Project administration: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Resources: Hollister W. Herhold, Steven R. Davis.

Software: Hollister W. Herhold.

Supervision: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Validation: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Visualization: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Writing – original draft: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

Writing – review & editing: Hollister W. Herhold, Steven R. Davis, David A. Grimaldi.

References

1. Weber RE, Vinogradov SN. Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol Rev. 2001; 81: 569–628. https://doi.org/10.1152/physrev.2001.81.2.569 PMID: 11274340
2. Burmester T. Evolution of Respiratory Proteins across the Pancrustacea. Integr Comp Biol. 2015; 55: 792–801. https://doi.org/10.1093/icb/icv079 PMID: 26130703
3. Wawrowski A, Matthews PGD, Gleixner E, Kiger L, Marden MC, Hankeln T, et al. Characterization of the hemoglobin of the backswimmer Anisops deanei (Hemiptera). Insect Biochem Mol Biol. 2012; 42: 603–609. https://doi.org/10.1016/j.ibmb.2012.04.007 PMID: 22575160
4. Hankeln T, Klawitter S, Krämer M, Burmester T. Molecular characterization of hemoglobin from the honeybee *Apis mellifera*. J Insect Physiol. 2006; 52: 701–710. https://doi.org/10.1016/j.jinsphys.2006.03.010 PMID: 16698031
5. Kuwada T, Hasegawa T, Sato S, Sato I, Ishikawa K, Takagi T, et al. Crystal structures of two hemoglobin components from the midge larva Propsilocerus akamusi (Orthocladiinae, Diptera). Gene. 2007; 398: 29–34. https://doi.org/10.1016/j.gene.2007.02.049 PMID: 17590288
6. Leitch I. The function of haemoglobin in invertebrates with special reference to *Planorbis* and *Chironomus* larvae. J Physiol. 1916; 50: 370–9. https://doi.org/10.1113/jphysiol.1916.sp001762 PMID: 16993350
7. Pause I. Beiträge zur Biologie und Physiologie der Larve von Chironomus gregarius. Zool Jahrb. 1918; 36: 1–114.
8. Scheer D. Die Farbstoffe der Chironomidenlarven. Arch Hydrobiol. 1934; 23: 391.
9. Keilin D, Wang YL. Haemoglobin of *Gastrophilus* larvae. Purification and properties. Biochem J. 1946; 40: 855–866. https://doi.org/10.1042/bj0400855 PMID: 16748097
10. Hungerford HB. Oxyhaemoglobin present in backswimmer Buenoa margaritacea Bueno (Hemiptera). Can Entomol. 1922; 54: 262–3.
11. Miller PL. Possible Function of Hæmoglobin in *Anisops*. Nature. 1964; 201: 1052. https://doi.org/10.1038/2011052a0 PMID: 14191597
12. Braunitzer G, Glossman H, Horst J. Zur Frage der nativen Hämoglobine der Larven von *Chironomus thummi thummi*. Hoppe-Seyler's Z Physiol Chem. 1968; 349: 1789–1791. PMID: 5707045
13. Bruan V, Crichton RR, Braunitzer G. Über monomere und dimere Insectenhämoglobine aus *Chironomus thummi*. Hoppe-Seyler's Z Physiol Chem. 1968; 349: 197–210. PMID: 4877820
14. Burmester T, Storf J, Hasenjäger A, Klawitter S, Hankeln T. The hemoglobin genes of *Drosophila*. FEBS J. 2006; 273: 468–480. https://doi.org/10.1111/j.1742-4658.2005.05073.x PMID: 16420471
15. Burmester T, Klawitter S, Hankeln T. Characterization of two globin genes from the malaria mosquito *Anopheles gambiae*: Divergent origin of nematoceran haemoglobins. Insect Mol Biol. 2007; 16: 133–142. https://doi.org/10.1111/j.1365-2583.2006.00706.x PMID: 17298561
16. Burmester T, Hankeln T. The respiratory proteins of insects. J Insect Physiol. 2007; 53: 285–294. https://doi.org/10.1016/j.jinsphys.2006.12.006 PMID: 17303160
17. English DS. Ontogenetic changes in hemoglobin synthesis of two strains of *Chironomus tentans*. J Embryol Exp Morphol. 1969; 22: 465–476. PMID: 5360026
18. Thompson P, Bleecker W, English DS. Molecular size and subunit structure of the hemoglobins of Chironomus tentans. J Biol Chem. 1968; 243: 4463–4467. PMID: 5684003
19. Tichy H. Hemoglobins of Chironomus tentans and C. pallidivittatus: Biochemical and Cytological Studies. Molecular Genetics. Berlin, Heidelberg: Springer; 1968. pp. 248–252.
20. Hankeln T, Jaenicke V, Kiger L, Dewilde S, Ungerechts G, Schmidt M, et al. Characterization of *Drosophila* hemoglobin. Evidence for hemoglobin-mediated respiration in insects. J Biol Chem. 2002; 277: 29012–29017. https://doi.org/10.1074/jbc.M204009200 PMID: 12048208
21. Burmester T, Hankeln T. Letter To The Editor: A Globin Gene of *Drosophila melanogaster*. Mol Biol Evol. 1999; 16: 1809–1811. https://doi.org/10.1093/oxfordjournals.molbev.a026093 PMID: 10605122
22. Dewilde S, Blaxter M, Van Hauwaert ML, Van Houte K, Pesce A, Griffon N, et al. Structural, functional, and genetic characterization of *Gastrophilus* hemoglobin. J Biol Chem. 1998; 273: 32467–32474. https://doi.org/10.1074/jbc.273.49.32467 PMID: 9829978
23. Pesce A, Nardini M, Dewilde S, Hoogewijs D, Ascenzi P, Moens L, et al. Modulation of oxygen binding to insect hemoglobins: The structure of hemoglobin from the botfly *Gasterophilus intestinalis*. Protein Sci. 2005; 14: 3057–3063. https://doi.org/10.1110/ps.051742605 PMID: 16260762
24. Bergtrom G. Partial characterization of haemoglobin of the bug, *Buenoa confusa*. Insect Biochem. 1977; 7: 313–316. https://doi.org/10.1016/0020-1790(77)90031-2

25. Osmulski PA, Vossbrinck CR, Sampath V, Caughey WS, Debrunner PG. Spectroscopic studies of an insect hemoglobin from the backswimmer Buenoa margaritacea (Hemiptera: Notonectidae). Biochem Biophys Res Commun. 1992; 87: 570–576.

26. Kawaoka S, Katsuma S, Meng Y, Hayashi N, Mita K, Shimada T. Identification and characterization of globin genes from two lepidopteran insects, Bombyx mori and Samia cynthia ricini. Gene. 2009; 431: 33–38. https://doi.org/10.1016/j.gene.2008.11.004 PMID: 19059317

27. Burmester T, Wawrowski A, Diepenbruck I, Schrick K, Seiwert N, Ripp F, et al. Divergent roles of the *Drosophila melanogaster* globins. J Insect Physiol. 2018; 106: 224–231. https://doi.org/10.1016/j.jinsphys.2017.06.003 PMID: 28606854

28. Yadav R, Sarkar S. *Drosophila* glob1 is required for the maintenance of cytoskeletal integrity during oogenesis. Dev Dyn. 2016; 245: 1048–1065. https://doi.org/10.1002/dvdy.24436 PMID: 27503269

29. Flogel U, Merx MW, Godecke A, Decking UKM, Schrader J. Myoglobin: A scavenger of bioactive NO. Proc Natl Acad Sci U S A. 2001; 98: 4276.

30. Freitas TAK, Saito JA, Hou S, Alam M. Globin-coupled sensors, protoglobins, and the last universal common ancestor. J Inorg Biochem. 2005; 99: 23–33. https://doi.org/10.1016/j.jinorgbio.2004.10.024 PMID: 15598488

31. Hetz SK, Bradley TJ. Insects breathe discontinuously to avoid oxygen toxicity. Nature. 2005; 433: 516–519. https://doi.org/10.1038/nature03106 PMID: 15690040

32. Franz-Guess S, Klußmann-Fricke BJ, Wirkner CS, Prendini L, Starck JM. Morphology of the tracheal system of camel spiders (Chelicerata: Solifugae) based on micro-CT and 3D-reconstruction in exemplar species from three families. Arthropod Struct Dev. 2016; 45: 440–451. https://doi.org/10.1016/j.asd.2016.08.004 PMID: 27519794

33. Kamentz C, Prendini L. An atlas of book lung ultrastructure in the Order Scorpiones (Arachnida). Bull Am Museum Nat Hist. 2008; 316: 1–259.

34. Schmidt C, Wägele JW. Morphology and evolution of respiratory structures in the pleopod exopodites of terrestrial isopoda (Crustacea, Isopoda, Oniscidea). Acta Zool. 2001; 82: 315–330. https://doi.org/10.1046/j.1463-6395.2001.00092.x

35. Giribet G, Edgecombe GD. The Phylogeny and Evolutionary History of Arthropods. Curr Biol. 2019; 29: R592–R602. https://doi.org/10.1016/j.cub.2019.04.057 PMID: 31211983

36. Dunlop JA, Anderson LI, Kerp H, Hass H. Preserved organs of Devonian harvestmen. Nature. 2003; 425: 916. https://doi.org/10.1038/425916a PMID: 14586459

37. Franz-Guess S, Starck JM. Histological and ultrastructural analysis of the respiratory tracheae of Galeodes granti (Chelicerata: Solifugae). Arthropod Struct Dev. 2016; 45: 452–461. https://doi.org/10.1016/j.asd.2016.08.003 PMID: 27531444

38. Ramírez MJ. Respiratory System Morphology and the Phylogeny of Haplogyne Spiders (Araneae, Araneomorphae). J Arachnol. 2000; 28: 149–157. https://doi.org/10.1636/0161-8202(2000)028[0149:rsmatp]2.0.co;2

39. Osmulski PA, Leyko W. Structure, function and physiological role of chironomus haemoglobin. Comp Biochem Physiol—Part B Biochem. 1986; 85: 701–722. https://doi.org/10.1016/0305-0491(86)90166-5

40. Wells RMG, Hudson MJ, Brittain T. Function of the hemoglobin and the gas bubble in the backswimmer Anisops assimilis (Hemiptera: Notonectidae). J Comp Physiol B. 1981; 142: 515–522. https://doi.org/10.1007/BF00688984

41. Hourdez S, Lamontagne J, Peterson P, Weber RE, Fisher CR. Hemoglobin from a deep-sea: Hydrothermal-vent copepod. Biol Bull. 2000; 199: 95–99. https://doi.org/10.2307/1542868 PMID: 11081707

42. Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science (80-). 2016; 354: 336–339. https://doi.org/10.1126/science.aaf9070 PMID: 27846568

43. Vinogradov SN, Hoogewijs D, Bailly X, Arredondo-Peter R, Guertin M, Gough J, et al. Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. Proc Natl Acad Sci. 2005; 102: 11385–11389. https://doi.org/10.1073/pnas.0502103102 PMID: 16061809

44. National Center for Biotechnology Information [Internet]. NCBI Transcriptome Shotgun Assembly Database. 2018 [cited 1 Jun 2018]. Available: https://www.ncbi.nlm.nih.gov/Traces/wgs/?term=tsa

45. Waterhouse AM, Procter JB, Martin DM., Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench Bioinformatics. Bioinformatics. 2009; 25: 1189–1191. https://doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

46. Bauernfeind AL, Babbitt CC. The predictive nature of transcript expression levels on protein expression in adult human brain. BMC Genomics. 2017; 18: 1–11. https://doi.org/10.1186/s12864-016-3406-7 PMID: 28049423

47. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012; 13: 227–232. https://doi.org/10.1038/nrg3185 PMID: 22411467

48. Straub L. Beyond the transcripts: What controls Protein Variation? PLoS Biol. 2011; 9: 9–10. https://doi.org/10.1371/journal.pbio.1001146 PMID: 21909242

49. Hankeln T, Amid C, Weich B, Niessing J, Schmidt ER. Molecular evolution of the globin gene cluster E in two distantly related midges, *Chironomus pallidivittatus* and *C. thummi thummi*. J Mol Evol. 1998; 46: 589–601. https://doi.org/10.1007/pl00006339 PMID: 9545469

50. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. Science (80-). 2014; 346: 763–767. https://doi.org/10.1126/science.1257570 PMID: 25378627

51. Gleixner E, Herlyn H, Zimmerling S, Burmester T, Hankeln T. Testes-specific hemoglobins in *Drosophila* evolved by a combination of sub- and neofunctionalization after gene duplication. BMC Evol Biol. 2012; 12: 34. https://doi.org/10.1186/1471-2148-12-34 PMID: 22429626

52. Blank M, Burmester T. Widespread occurrence of N-terminal acylation in animal globins and possible origin of respiratory globins from a membrane-bound ancestor. Mol Biol Evol. 2012; 29: 3553–3561. https://doi.org/10.1093/molbev/mss164 PMID: 22718912

53. Hoogewijs D, Ebner B, Germani F, Hoffmann FG, Fabrizius A, Moens L, et al. Androglobin: A chimeric globin in metazoans that is preferentially expressed in mammalian testes. Mol Biol Evol. 2012; 29: 1105–1114. https://doi.org/10.1093/molbev/msr246 PMID: 22115833

54. Hoffmann FG, Opazo JC, Storz JF. Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates. Mol Biol Evol. 2012; 29: 303–312. https://doi.org/10.1093/molbev/msr207 PMID: 21965344

55. Hoffmann FG, Opazo JC, Hoogewijs D, Hankeln T, Ebner B, Vinogradov SN, et al. Evolution of the globin gene family in deuterostomes: Lineage-specific patterns of diversification and attrition. Mol Biol Evol. 2012; 29: 1735–1745. https://doi.org/10.1093/molbev/mss018 PMID: 22319164

56. Lechauve C, Jager M, Laguerre L, Kiger L, Correc G, Leroux C, et al. Neuroglobins, pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity. J Biol Chem. 2013; 288: 6957–6967. https://doi.org/10.1074/jbc.M112.407601 PMID: 23288852

57. Python Software Foundation. Python Language Reference, Versions 3.6 and 3.7. 2019.

58. R Core Team. R: A language and environment for statistical computing, Version 3.6.3. Vienna, Austria; 2013. Available: http://www.r-project.org

59. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25: 1422–1423. https://doi.org/10.1093/bioinformatics/btp163 PMID: 19304878

60. R Studio Team. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.; 2015. Available: http://www.rstudio.com

61. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

62. National Center for Biotechnology Information [Internet]. Non-redundant (nr) database. 2019 [cited 1 Jun 2018]. Available: http://ftp.ncbi.nlm.nih.gov/blast/db/

63. Larsson A. AliView: A fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014; 30: 3276–3278. https://doi.org/10.1093/bioinformatics/btu531 PMID: 25095880

64. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30: 3059–66. Available: http://www.ncbi.nlm.nih.gov/pubmed/12136088%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC135756 https://doi.org/10.1093/nar/gkf436 PMID: 12136088

65. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol. 2016; 33: 1635–1638. https://doi.org/10.1093/molbev/msw046 PMID: 26921390

66. Rasmus W, Pedersen AG. RevTrans—Constructing alignments of coding DNA from aligned amino acid sequences. Nucl Acids Res. 2003; 31: 3537–3539. https://doi.org/10.1093/nar/gkg609 PMID: 12824361

67. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

68. Jones DT. Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. J Mol Biol. 1999; 195–202. https://doi.org/10.1006/jmbi.1999.3091 PMID: 10493868

69. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015; 31: 926–932. https://doi.org/10.1093/bioinformatics/btu739 PMID: 25398609

70. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008; 9: 40. https://doi.org/10.1186/1471-2105-9-40 PMID: 18215316

71. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc. 2010; 5: 725–738. https://doi.org/10.1038/nprot.2010.5 PMID: 20360767

72. Zhang Y, Skolnick J. TM-align: A protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005; 33: 2302–2309. https://doi.org/10.1093/nar/gki524 PMID: 15849316

73. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics. 2017; 33: 3387–3395. https://doi.org/10.1093/bioinformatics/btx431 PMID: 29036616

74. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020; 48: D265–D268. https://doi.org/10.1093/nar/gkz991 PMID: 31777944

75. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.51. 2018.

76. Project Jupyter. JupyterLab. Project Jupyter; 2019. Available: http://www.jupyter.org

77. Jones E, Oliphant T, Peterson P, others. SciPy: Open source scientific tools for Python. Available: http://www.scipy.org/

78. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, et al. Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A. 2011; 108: 5690–5. https://doi.org/10.1073/pnas.1012675108 PMID: 21402926

79. Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, et al. Evolutionary History of the Hymenoptera. Curr Biol. 2017; 27: 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027 PMID: 28343967

80. Mitter C, Davis DR, Cummings MP. Phylogeny and Evolution of Lepidoptera. Annu Rev Entomol. 2017; 62: 265–283. https://doi.org/10.1146/annurev-ento-031616-035125 PMID: 27860521