THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 276, No. 28, Issue of July 13, pp. 26391–26397, 2001
Printed in U.S.A.

The Evolution of Extracellular Hemoglobins of Annelids, Vestimentiferans, and Pogonophorans*

Received for publication, January 19, 2001, and in revised form, April 2, 2001
Published, JBC Papers in Press, April 6, 2001, DOI 10.1074/jbc.M100557200

Enrico Negrisolo‡§, Alberto Pallavicini‡§, Roberto Barbato¶, Sylvia Dewilde‖, Anna Ghiretti-Magaldi‡‡, Luc Moens‖, and Gerolamo Lanfranchi‡§§**

From the ‡Dipartimento di Biologia and Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative, Università di Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy, the ¶Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale Amedeo Avogadro, Corso Borsalino 54 I-15100 Alessandria, Italy, and the ‡‡Dipartimento di Biologia and Consiglio Nazionale delle Ricerche Center for the Study of Metalloproteins, Università di Padova, via Ugo Bassi 58/B, 35131 Padova, Italy, and the ‖Department of Biochemistry, University of Antwerp UIA, B-2610 Wilrijk, Belgium

The evolution of extracellular hemoglobins of annelids, vestimentiferans, and pogonophorans was investigated by applying cladistic and distance-based approaches to reconstruct the phylogenetic relationships of this group of respiratory pigments. We performed this study using the aligned sequences of globin and linker chains that are the constituents of these complex molecules. Three novel globin and two novel linker chains of *Sabella spallanzanii* described in an accompanying paper (Pallavicini, A., Negrisolo, E., Barbato, R., Dewilde, S., Ghiretti-Magaldi, A., Moens, L., and Lanfranchi, G. (2001) *J. Biol. Chem.* 276, 26384–26390) were also included. Our results allowed us to test previous hypotheses on the evolutionary pathways of these proteins and to formulate a new most parsimonious model of molecular evolution. According to this novel model, the genes coding for the polypeptides forming these composite molecules were already present in the common ancestor of annelids, vestimentiferans, and pogonophorans.

The extracellular hemoglobins of Annelida, Vestimentifera, and Pogonophora together with the chlorocruorin, a variant extracellular hemoglobin restricted to four polychaete families, have been recognized to form a unique group according to their physicochemical properties (1). These molecules are complexes of a large number of polypeptide components that assemble into hierarchical ordered quaternary structures.

The extracellular Hbs and Chls[1] of Annelida are formed by two kinds of polypeptide chains, the globins and the linkers (2). Vestimentifera have two types of extracellular Hbs (2) that are indicated here as heavy hemoglobin (Hb_H) and light hemoglobin (Hb_L). Hb_H is very similar in shape and molecular mass (~ 3,000 kDa) to the Hbs and Chls of Annelida and is also composed of globin and linker chains. In contrast, the Hb_L has a molecular mass of ~ 400 kDa and contains globin chains only. Pogonophora contain an Hb very similar to the Hb_L of Vestimentifera; therefore, it is also indicated here as Hb_L (2).

In their pioneering work on the molecular evolution of extracellular Hbs, Gotoh *et al.* (3) divided the globin polypeptides into two groups with a common origin. Successively, Suzuki and Riggs (4) demonstrated that linker chains share similarities with the low density lipoprotein receptor and that they cannot be related to the globins because of the high divergence of their primary structures. More recently, Yuasa *et al.* (5) have proposed a general evolutionary model for the Hbs, Hb_H, and Hb_L. According to these authors the common ancestor of all these molecules was a protein formed by globin chains only. During the successive evolution the linker chains would have been added to the final structure of Hbs and Hb_H. Chls were not included in their analysis, because no sequences were available at that time. Moreover, their model implies an independent evolution of Hbs and Hb_H. As a consequence two points still remain unsettled: (i) the placement of Chl in this scenario and (ii) the evolutionary pathway followed by Hbs, Chls, and Hb_H with respect to Hb_L.

To investigate these points we have sequenced the globin and linker chains of Chl of the polychaete *Sabella spallanzanii* (6). They have been aligned with other available sequences of Annelida, Vestimentifera, and Pogonophora to perform phylogenetic analyses applying both cladistic and distance-based methods. Using this approach we have been able to retrace the molecular evolution of Hbs, Chls, Hb_Hs, and Hb_Ls. The comparison of our new phylogenetic reconstruction with the previously advanced hypotheses (3, 5) lead us to propose an alternative model for the evolution of annelid hemoglobins.

MATERIALS AND METHODS

Sequence Alignment

We aligned all the sequences of extracellular Hbs, Chls, Hb_Hs, and Hb_Ls of Annelida, Pogonophora, and Vestimentifera available in the literature including the newly presented globin and linker sequences of *S. spallanzanii* (6). Globin sequences were aligned manually according to the nonvertebrate globin template (7), and linker sequences were aligned using the ClustalW program (8).

Pairwise Comparisons

Pairwise comparisons were performed on the coding and 3′-UTR portions of the *S. spallanzanii* globin cDNAs and some selected globins of Annelida using the program ALIGN. Similar comparisons were carried out on the linker sequences using the ClustalW program (8).

* This work was supported in part by the Italian Ministry of University and of Scientific and Technological Research MURST Grant-Cofinanziamento Protocol 9805192993-002 (to A. G.-M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Both authors contributed equally to this work.

** Post-doctoral fellow of the Fund for Scientific Research-Flanders (FWO).

§§ To whom correspondence should be addressed. Tel.: 39-04982-76221; Fax: 39-04982-76280; E-mail: lanfra@cribi.unipd.it.

[1] The abbreviations used are: Chl, chlorocruorin; Hb_H, heavy hemoglobin; Hb_L, light hemoglobin; UTR, untranslated region; NJ, neighbor-joining; SC, strict consensus.

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

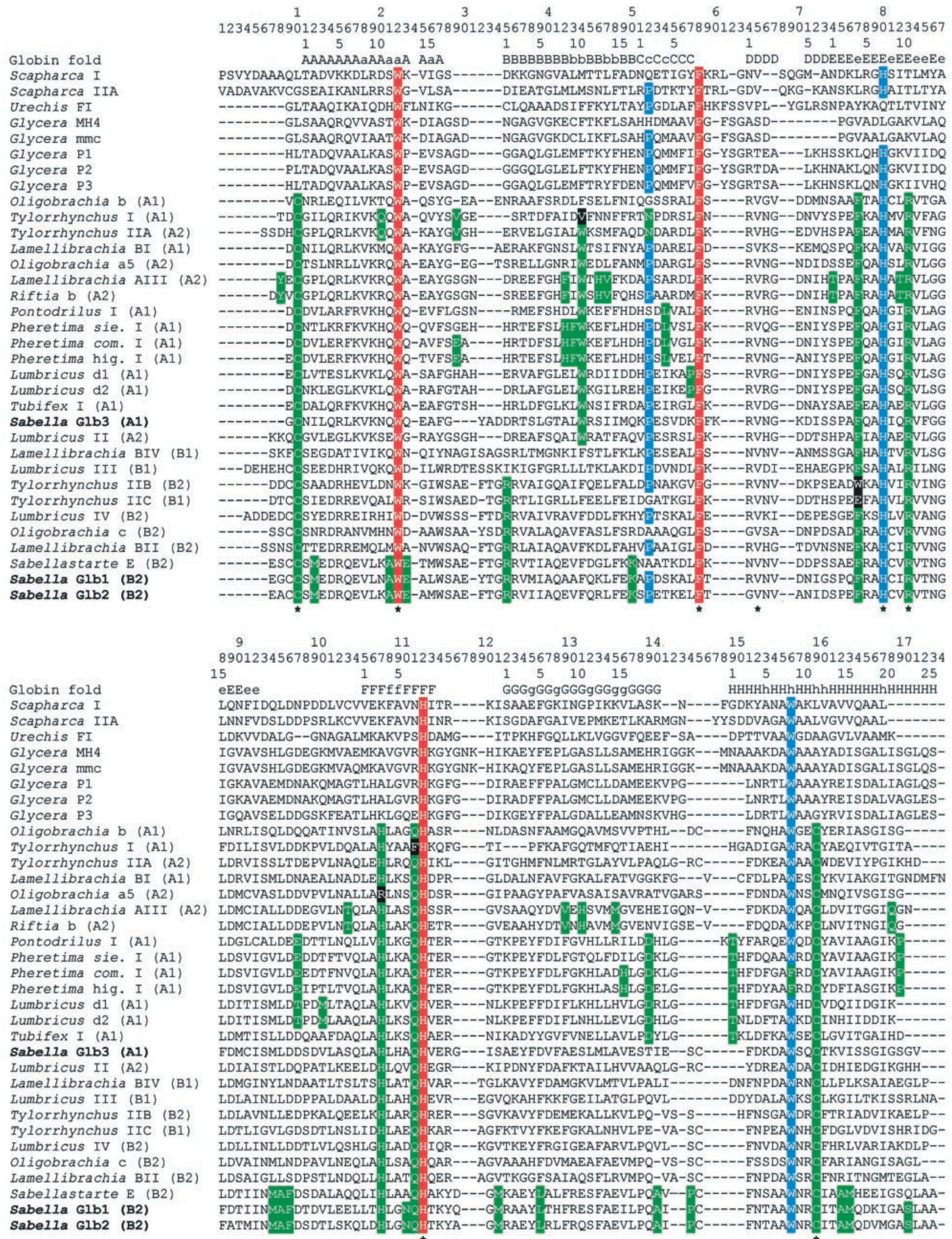

FIG. 1. **Alignment of globin chains.** The sequences have been aligned in the coding portions using as reference the tertiary structure template of invertebrate globins (7). They are shown with the accession numbers for the Swiss-Prot, NCBI, and EMBL data banks in parentheses: *Scapharca* I (P02213), *S. inaequivalvis*; *Scapharca* IIA (P14821), *S. inaequivalvis*; *Urechis* F1 (P06148), *U. caupo*; *Glycera* MH4 (P15447), *G. dibranchiata*; *Glycera* mmc (P02216), *G. dibranchiata*; *Glycera* I (P23216), *G. dibranchiata*; *Glycera* II (P21659), *G. dibranchiata*; *Glycera* III (P21660), *G. dibranchiata*; *Oligobrachia* b (5), *Oligobrachia mashikoi*; *Tylorrhynchus* I (P02219), *Tylorrhynchus heterochaetus*; *Tylorrhynchus* IIA (P09966), *Tylorrhynchus heterochaetus*; *Lamellibrachia* BI (23), *Lamellibrachia* sp.; *Oligobrachia* a5 (5), *Oligobrachia mashikoi*; *Lamellibrachia* AIII (P15469), *Lamellibrachia* sp.; *Riftia* b (P80592), *Riftia pachyptila*; *Pontodrilus* I (24), *Pontodrilus matsushimensis*; *Pheretima sie*. I (P11740), *Pheretima sieboldi*; *Pheretima com*. I (24), *Pheretima comunissima*; *Pheretima hil*. I (25), *Pheretima hilgendorfi*; *Lumbricus* d1 (U55073),

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

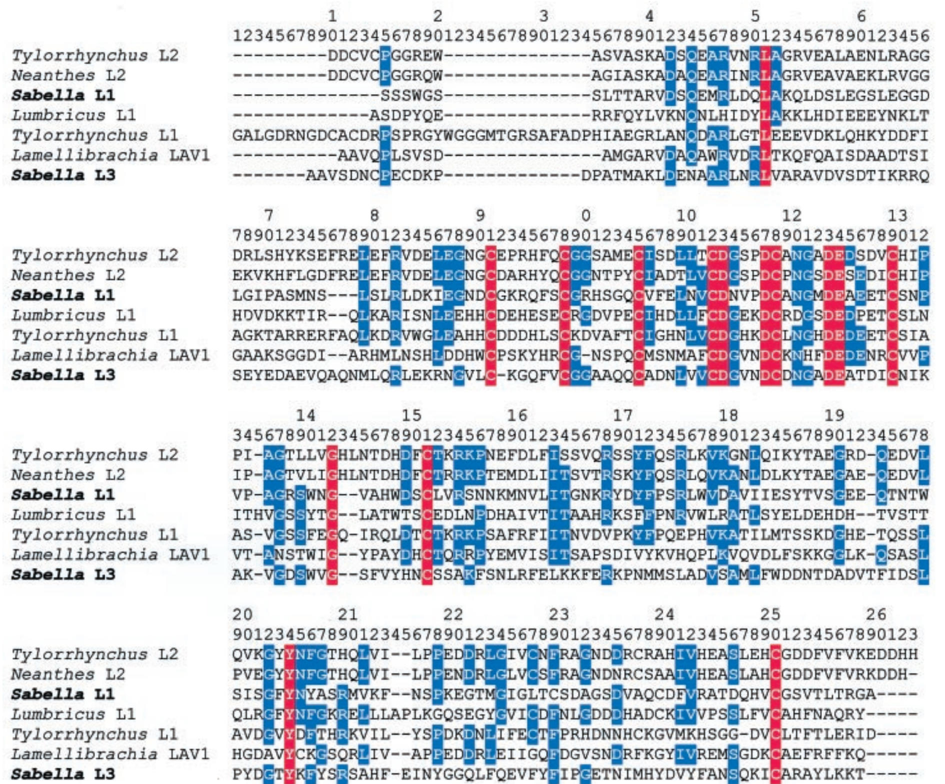

FIG. 2. **Alignment of linker chains.** Amino acid sequences of the following species were used (with the accession numbers from Swiss-Prot, NCBI, and EMBL data banks in parenthesis): *Tylorrhynchus* L2 (P18208), *Tylorrhynchus heterochaetus*; *Neanthes* L2 (D58413), *Neanthes diversicolor*; *Sabella* L1 (AJ131900), *S. spallanzanii*; *Lumbricus* L1 (A46587), *Lumbricus terrestris*; *Tylorrhynchus* L1 (P18207), *Tylorrhynchus heterochaetus*; *Lamellibrachia* LAV1 (P16222), *Lamellibrachia* sp.; and *Sabella* L3 (AJ131286), *S. spallanzanii*. *Red*, invariant amino acids; *blue*, amino acids common to the majority of sequences.

Phylogenetic Analyses

Sequence alignments were used to build phylogenetic trees using the maximum parsimony[2] (9) and neighbor-joining (NJ) methods (10).

Outgroup Choice—Globin sequences with known x-ray structures were chosen as outgroups. Based on the phylogeny of Metazoa (11), the sequences of the closest relatives of Annelida, Vestimentifera, and Pogonophora were selected as outgroups. These groups are (i) the globins of the mollusk *Scapharca inaequivalvis*, (ii) the globin of the echiuran *Urechis caupo*, and (iii) the intracellular globins of the annelid *Glycera dibranchiata*. No outgroup for the linker sequences could be selected because their evolutionary position is unknown (4).

Cladistic Analysis—The cladistic analyses were done using the program PAUP, version 3.1.1. (12). Globin evolution was studied using a heuristic approach because the number of sequences involved excludes the exhaustive approach (12). All characters were weighted equally, and those uninformative were excluded from the analysis. After a series of trials, we set the PAUP options to perform more efficiently in finding the shortest trees. For the optimization of characters, the *ACCTRAN* option was set. For the heuristic searches, the options set were: (i) *keep minimal trees only*, (ii) *collapse zero-length branches*, (iii) starting tree sources, *get by stepwise addition*, (iv) swap on, *minimal trees only*, (v) addition sequence, *random*, (vi) replications, *50*, (vii) swapping algorithm, *TBR tree bisection-reconnection*, (viii) *MULPARS on*, and (ix) *steepest descent on*.

For the phylogenetic analysis of linker chains, we applied an exhaustive approach because of the smaller size of the data set (12). Uninformative characters were also excluded, and the remaining characters were weighted equally. The PAUP optimization of characters (*ACCTRAN on*) and exhaustive search (*keep minimal trees only* and *collapse zero-length branches on*) options were set.

Tree Indexes in the Cladistic Analyses—Two indexes are traditionally used to test the robustness of the most parsimonious tree obtained by cladistic analyses, the consistency and retention indexes (13, 14). These indexes were calculated as implemented in PAUP (12). Values of the consistency index and retention index > 0.5 indicate that convergent/parallel evolution does not affect strongly the phylogenetic reconstruction and that the obtained topologies of trees are reliable. Molecular synapomorphies in the cladistic analysis were detected using the program MacClade (Version 3), which allows the tracking of the evolution of each character (15).

Neighbor-joining Analysis—The distance-based phylogenetic reconstructions were performed on globin and linker alignments according to the NJ method (10) as implemented in the TREECON program (Version

[2] Glossary of phylogenetic terms: apomorphy, a derived state of a character that represents an evolutionary novelty with respect to the ancestral state; autoapomorphy, a derived character state that is unique for a particular sequence; bootstrap test, a statistical test used to verify the robustness of the topology of an evolutionary tree; bootstrap value, the result of the bootstrap test (in this paper it is expressed as a percentage); cladistic analysis, a phylogenetic reconstruction based on the principle of maximum parsimony; cladogram, a tree depicting the phylogenetic relationships that results from a cladistic approach; clade, a monophyletic group in a cladistic context; ingroup, a set of sequences that are considered the focus of interest; monophyletic group, a group that includes all the sequences that originate from a common ancestor; most parsimonious tree, a tree produced in a cladistic analysis that has the shortest length; orthologous sequences, two sequences derived from a speciation event (*i.e.* the same sequence in different species); outgroups, a set of sequences that are brought into the analysis to determine the root of the ingroup and ancestral states; strict consensus tree, a tree derived from a set of trees in which all conflicting branching patterns are collapsed into multifurcations; synapomorphy, a derived state of a character that is shared by all the taxa belonging to a clade; tree length in the cladistic analysis, the value obtained by computing the sum of the minimum numbers of substitutions in all the positions of the alignment.

Lumbricus terrestris; *Lumbricus* d2 (U55074), *Lumbricus terrestris*; *Tubifex* 1 (P18202), *Tubifex tubifex*; *Sabella* Glb3 (AJ131285), *S. spallanzanii*; *Lumbricus* II (P02218), *Lumbricus terrestris*; *Lamellibrachia* BIV (23), *Lamellibrachia* sp.; *Lumbricus* III (P11069), *Lumbricus terrestris*; *Tylorrhynchus* IIB (P13578), *Tylorrhynchus heterochaetus*; *Tylorrhynchus* IIC (P02220), *Tylorrhynchus heterochaetus*; *Lumbricus* IV (P13579), *Lumbricus terrestris*; *Oligobrachia* c (5), *Oligobrachia mashikoi*; *Lamellibrachia* BII (23), *Lamellibrachia* sp.; *Sabellastarte* E (D58418), *Sabellastarte indica*; *Sabella* Glb1 (AJ131283), *S. spallanzanii*; and *Sabella* Glb2 (AJ131284), *S. spallanzanii*. The amino acids common among all the globins of Annelida, Pogonophora, and Vestimentifera are indicated by *asterisks*. *Red*, invariant amino acids; *blue*, amino acids common to the majority of sequences; *green*, synapomorphic amino acids; *black*, autoapomorphic amino acids.

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

1.3b) (16). The NJ trees were created by applying the following settings: (i) distance calculation, *Kimura 83*; (ii) alignment positions, *all*; and (iii) insertions and deletions, *not taken into account*.

Bootstrap Test—The bootstrap resampling (17) was performed to test the robustness of the trees obtained by cladistic and neighbor-joining phylogenetic reconstructions. In both cases, 500 replicates were run.

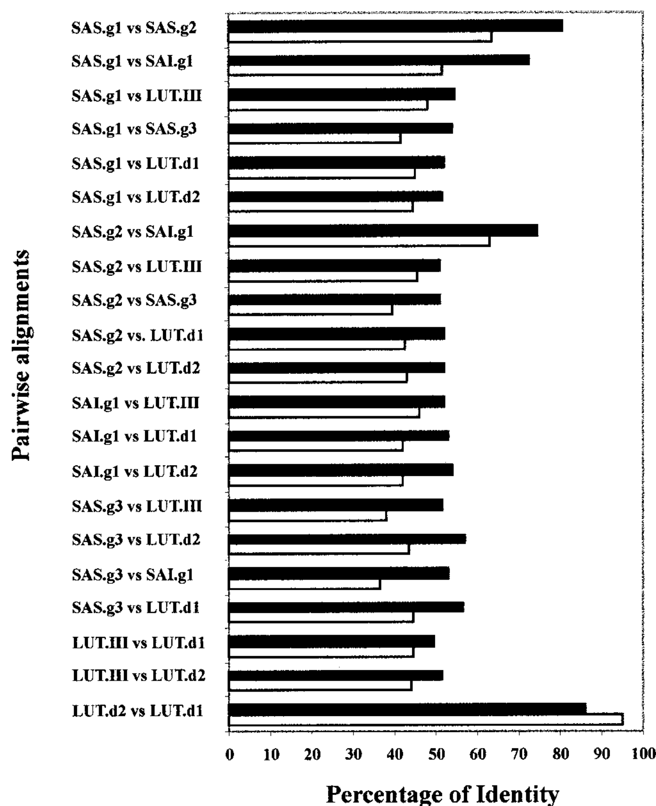

FIG. 3. **Pairwise comparison of open reading frames and 3′ UTRs of globin cDNAs.** *Black bars*, globin open reading frame; *white bars*, globin 3′ UTRs; *SAS.g1*, *Sabella* Glb1 globin; *SAS.g2*, *Sabella* Glb2 globin; *SAS.g3*, *Sabella* Glb3 globin; *SAI.gl*, *Sabellastarte* E globin; *LUT.d1*, *Lumbricus* d1 globin; *LUT.d2*, *Lumbricus* d2 globin; *LUT.III*, *Lumbricus* III globin.

RESULTS

Sequence Alignments

Globin Alignment—In Fig. 1 we present the results of a global alignment of the *Sabella* globin sequences with the globins of other annelids, vestimentiferans, pogonophorans, and the canonical globin fold (7). Eight sequences chosen according to the criteria described under "Materials and Methods" were added as outgroups. An inspection of the alignment clearly indicates that the key residues Trp (A12), Phe (CD1), His (E7), and His (F8), consistent with the canonical globin fold, are absolutely conserved. Conversely, the presence of Pro (C2), which usually determines the bend of the corner between helices B and C, is not universal to all the globins. In two *Sabella* globins (Glb1 and Glb2), the small residues Ala and Cys are found at positions B6 and E8. In these sites the majority of annelid globins have a Gly residue, with the exception of globin I of *Tylorrhynchus*, in which a Phe is found at position B6. These substitutions in the *Sabella* globins would result in a closer crossing of the B and E helices. All the extracellular annelid globins shown in the alignment have in common the Cys residues at positions NA2 and H11. These amino acids are important for the formation of the first supramolecular aggregate in the assembly of the whole Hb (18), which is confirmed by the fact that they are replaced in the intracellular globins of *Glycera*, *Urechis*, and *Scapharca* by other residues. Four other residues, namely Arg (E10), Phe (E4), His (F3), and Gln (F7), are highly conserved in the extracellular globins, but the structural and functional significance of this conservation remains to be clarified.

Linker Alignment—A similar alignment was carried out for the linker sequences of *Sabella* with the five other linkers available in the data bases (Fig. 2). The main feature that seems to be conserved is a cysteine-rich segment ((Cys-X_6)$_3$-Cys-X_5-Cys-X_{10}-Cys) that is typical for all the linker chains sequenced thus far and has been related to the low density lipoprotein-receptor motif (4). Considering the current global

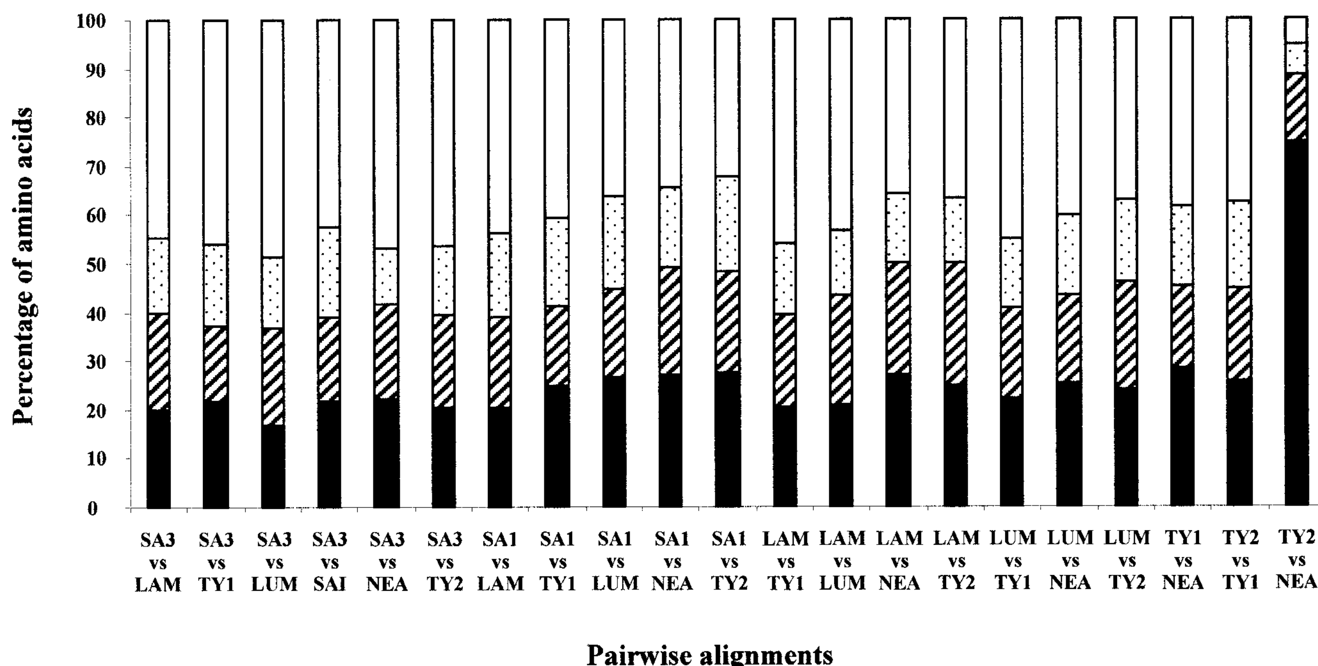

FIG. 4. **Pairwise comparison of amino acid sequences of linker chains.** *Black bars*, identical amino acids; *striped bars*, strongly similar amino acids; *dotted bars*, weakly similar amino acids; *white bars*, different amino acids; *TY2*, *Tylorrhynchus* L2 linker; *NEA*, *Neanthes* L2 linker; *SA1*, *Sabella* L1 linker; *LUM*, *Lumbricus* L1 linker; *TY1*, *Tylorrhynchus* L1 linker; *LAM*, *Lamellibrachia* LAV1 linker; *SA3*, *Sabella* L3 linker.

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

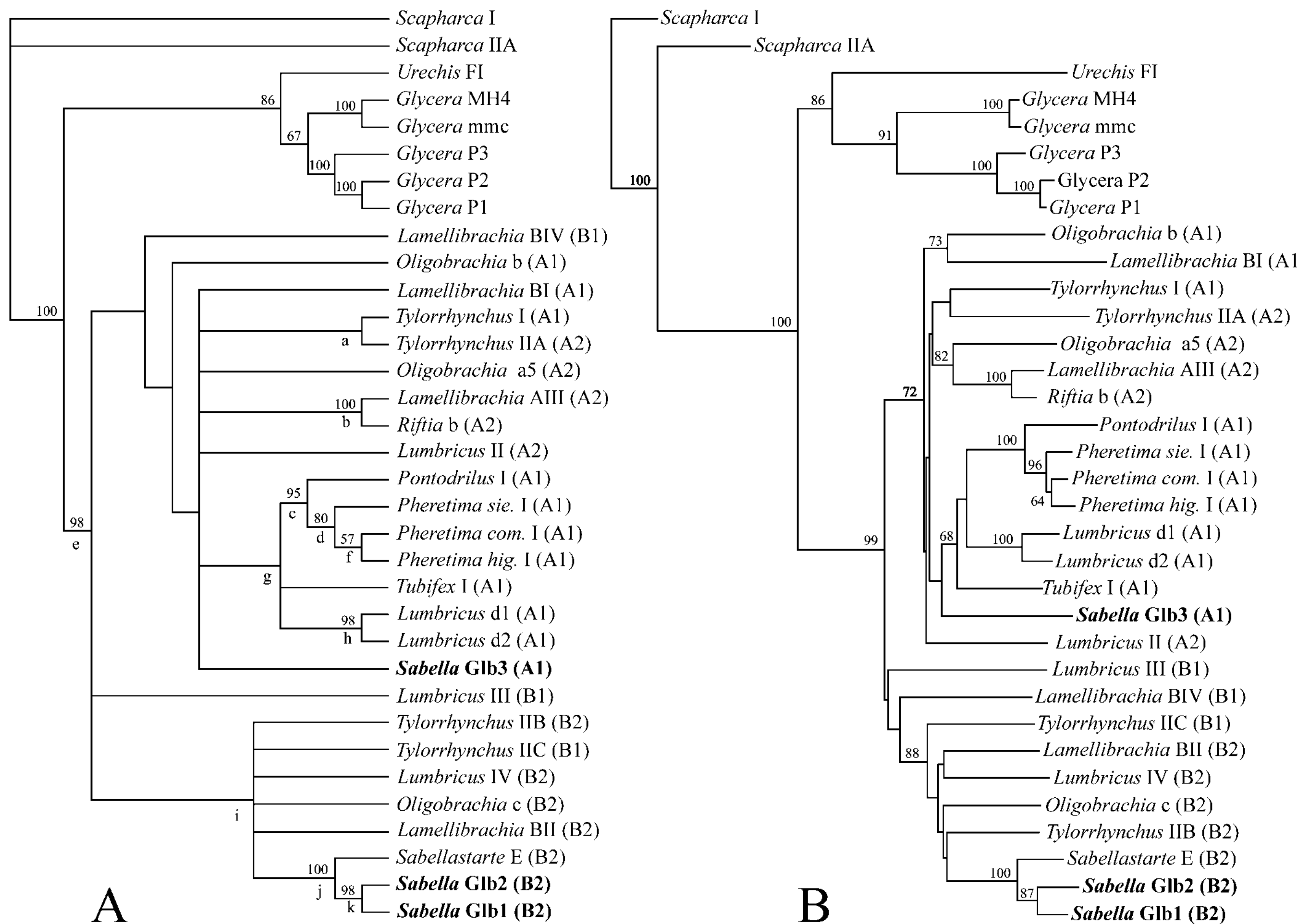

FIG. 5. **Phylogenetic analysis of globin polypeptides.** *A*, the SC tree derived from merging 18 most parsimonious cladograms. The statistic analyses of the most parsimonious cladograms gave the following values: length, 1,689 steps; consistency index, 0.601; retention index, 0.587. *B*, NJ tree. Both trees are based on the alignment shown in Fig. 1. The *numbers* on the branches refer to the bootstrap values expressed as a percentage after 500 replicates; only values ≥ 50% are reported. *Lowercase letters* located close to the nodes refer to monophyletic clades that are supported by the molecular synapomorphies reported in Table I. The standard abbreviations of globins are shown in *parentheses* according to the nomenclature previously suggested by Gotoh *et al.* (22).

alignment (Fig. 2), the cysteine-rich segment can be written more precisely as ((Cys-X_{5-6})$_2$-Cys-X_6-Cys-Asp-X_3-Asp-Cys-X_4-Asp-Glu-X_4-Cys).

Pairwise Comparison—The analysis of the pairwise comparison of the globin cDNAs (Fig. 3) shows a higher percentage of identity in the open reading frame portion (57.59% mean value) than in the 3′-UTR portion (47.82% mean value). This is consistent with a higher degree of variability of the globin 3′ UTRs that are less subject to structural constrains. However, it should be noted that *Lumbricus* d1 globin and d2 are more similar in 3′ regions than in the coding portions.

The pairwise comparisons performed on the linker chains (Fig. 4) give the following main results. (i) *Tylorrhynchus* L2 and *Neanthes* L2 markedly differ from the general structure of other linkers. In fact they share 74.58% of identical amino acids, whereas the mean for the whole data set is 25.79%. The latter value decreases to 23.35%, if the comparison of *Tylorrhynchus* L2 *versus Neanthes* L2 is excluded from the computation. (ii) Roughly 40% of the amino acids are specific for each chain (mean of dissimilar amino acids = 39.45%). (iii) Identical amino acids represent one fourth of the whole data when sequences are pairwise-compared. However, the residues common to all the linker chains in the global alignment are only 15 (5.7%) (Fig. 2), and they are mainly restricted to the cysteine-rich segment.

Phylogenetic Analyses

Globin Phylogenetic Analysis—We have applied both cladistic and distance-based methods to study the molecular evolution of annelid globins. The cladistic analysis results in 18 equally parsimonious cladograms. The resulting strict consensus (SC) tree is presented in Fig. 5*A* together with the distance-based tree (Fig. 5*B*). Both phylogenetic reconstructions recognize a monophyletic origin of Hbs, Chls, Hb_Hs, and Hb_Ls. A more detailed analysis of the ingroup shows that the extracellular Hbs, Chls, Hb_Hs, and Hb_Ls can be divided into two distinct groups. However, the actual position of the *Lumbricus* III globin cannot be resolved in the SC tree. The topologies of the two trees show some discrepancies. In particular the placement of the *Lamellibrachia* BIV globin seems controversial. In fact, in the cladogram it is placed into the A group according to the classification of Gotoh *et al.* (3), whereas in the NJ tree it is included in the B group. Both hypotheses are poorly supported by bootstrap values, revealing the weakness of the more basal nodes. In Table I we list the clades that belong to the ingroup and are supported by one or more molecular synapomorphies. Several of them are also sustained by high bootstrap values.

Linker Phylogenetic Analysis—The cladistic analysis of linker sequences produced two equally parsimonious cladograms. The SC tree is showed in Fig. 6, where the NJ tree is

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

TABLE I
Globin clades

The globin sequences aligned in Fig. 1 have been grouped according to the maximum parsimony criterion, and the common amino acids that represent the molecular synapomorphies for the different clades are listed together with their relative positions in the alignment. i, clade positions on the trees reported in Fig. 5. BT, bootstrap values supporting the different clades. Only values ≥ 50% are indicated.

i	Clade	BT	Molecular synapomorphies
		%	
a	*Tylorrhynchus* IIA + *Tylorrhynchus* I		Gln (20), Val (29), Asn (52)
b	*Lamellibrachia* AIII + *Riftia* b	100	Tyr (8), Phe (42), His (46), Val (47), Thr (74), Thr (82), Thr (103), Val (129), His (131), Met (135), Gln (168)
c	*Pontodrilus* I + (*Pheretima sie*. I + (*Pheretima com*. I + *Pheretima hig*. I))	95	Leu (54), Glu (97), Pro (169)
d	*Pheretima sie*. I + (*Pheretima com*. I + *Pheretima hig*. I)	96	His (42), Phe (43)
e	Hbs + Chls + Hb_Hs + Hb_Ls	98	Cys (10), Arg (83), Cys (159)
f	*Pheretima com*. I + *Pheretima hig*. I	68	Glu (29), His (136), Phe (156)
g	*Pontodrilus* I to *Lumbricus* d2;		Asp (139), Thr (149)
h	*Lumbricus* d1 + *Lumbricus* d2	98	Pro (57), Thr (97), Met (100)
i	*Tylorrhynchus* IIB to *Sabella* Glb1		Ser (35)
j	*Sabellastarte* E + (*Sabella* Glb2 + *Sabella* Glb1)	100	Met (12), Ala (21), Glu (23), Lys (50), Ala (95), Phe (96), Met (121), Leu (126), Ala (140), Pro (144), Ala (162), Met (163)
k	*Sabella* Glb2 + *Sabella* Glb1	98	Asn (110), Ser (170)

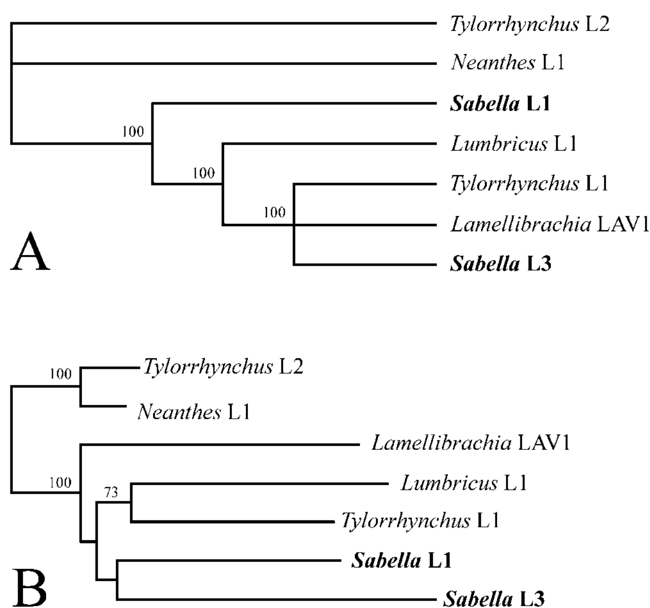

FIG. 6. **Phylogenetic analyses of linker polypeptides.** *A*, the SC tree derived from merging two most parsimonious cladograms The statistic analyses of the most parsimonious cladograms gave the following values: length, 326 steps; consistency index, 0.862; retention index, 0.612. *B*, NJ tree. Both trees are based on the alignment reported in Fig. 2. *Numbers* on the branches refer to the bootstrap values expressed as a percentage after 500 replicates; only values ≥ 50% are reported.

also presented. Cladistic and distance-based analyses were performed only on the available sequences of linker chains because no convincing putative outgroups are known. In fact, with the exception of the cysteine-rich segment that relates the linkers with the low density lipoprotein receptor (4), no other significant similarity is known between linker chains and other proteins. Nevertheless, *Tylorrhynchus* L2 and *Neanthes* L1 were used to root the trees, because they are much more similar to each other than to other linkers.

The SC and NJ trees show some discrepancies in their topology. The cladogram structure favors a strict relationship between *Tylorrhynchus* L1, *Lamellibrachia* LAV1, and *Sabella* L3. This clade is also supported by a very high bootstrap value. Conversely, the NJ tree supports the two groups *Tylorrhynchus* L1 + *Lumbricus* L1 and *Sabella* L1 + *Sabella* L3. The first is also corroborated by the bootstrap value (72%), whereas the second does not receive strong support by the bootstrap test (48%). Both cladistic and distance-based analyses strongly favor the grouping of *Tylorrhynchus* L2 and *Neanthes* L2 with respect to other linker chains.

DISCUSSION

The cladistic and distance-based analyses that we performed on the extracellular Hbs of annelids, pogonophorans, and vestimentiferans confirm the previous results of Gotoh *et al.* (3). These authors proposed the division of the globin chains into two main groups, A and B, each divided further into two subgroups, A1/A2 and B1/B2. This classification was successively applied to Hb_Hs and Hb_Ls (19, 20). Recently, a correction of this nomenclature has been proposed that recommends an inversion of the names of the two main strains (21).

However the identification of homologous chains, *i.e.* globins that are the products of orthologous genes, does not seem to be a trivial task. In both our analyses *Tylorrhynchus* IIA and *Tylorrhynchus* I are grouped together more closely than the respective "homologous" sequences. Therefore it seems problematic to name them *Tylorrhynchus* A2 and *Tylorrhynchus* A1 as previously suggested (22). In this light *Lumbricus* d1 and *Lumbricus* d2 can be defined either as two allelic forms of the same gene or as the products of a very recent gene duplication. The same reasoning can be applied to *Sabella* Glb1 and *Sabella* Glb2.

The subdivision into four homologous groups of globins could be an oversimplification of the real situation. In fact we have found a higher number of globin chains in the Chl, purified from a single specimen of *Sabella* (6). A study of the entire portion of the genome coding for these proteins should be the best way to understand these discrepancies.

The globins and linkers, forming the Chls, are tightly associated in the phylogenetic reconstructions with those included in the Hbs. This clearly identifies the Chl as a variant of Hb. On this assumption the name chlorocruorin must be considered only as a descriptive term.

The phylogenetic reconstruction presented in this paper reveals that the genes coding for the polypeptides that form Chls, Hbs, Hb_Hs, and Hb_Ls appeared before the separation of Vestimentifera and Pogonophora phyla from Annelida (*Sensu stricto*) (11). The evolutionary pathway does not change even if we consider pogonophorans and vestimentiferans as members of the class Opisthochaeta within the phylum Annelida (20). The critical question is: what was the scenario in which Chls, Hbs, Hb_Hs, and Hb_Ls evolved?

Yuasa *et al.* (5) previously suggested that Hb, Chl, and Hb_H evolved from Hb_L, which is made only of globin chains, adding linker chains to form the final hexagonal bilayered structure. If

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

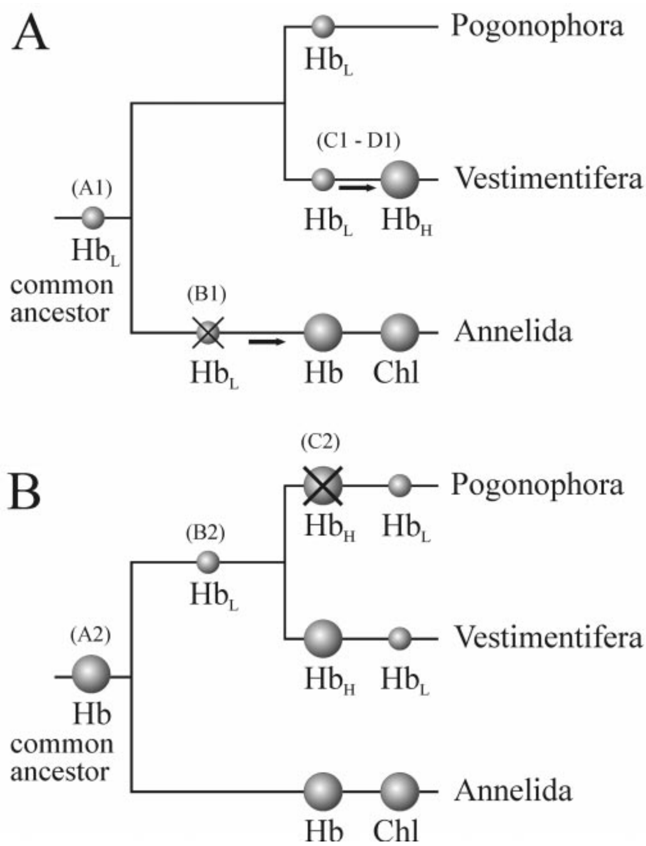

FIG. 7. **Evolutionary models of extracellular hemoglobins.** The two alternative hypotheses for the evolution of extracellular hemoglobins are shown. *Letters in parentheses* correspond to the different assumptions that support the two different models. See the text for discussion.

we accept this hypothesis we must make four assumptions (Fig. 7*A*): (i) Hb_L was present in the common ancestor of Annelida, Vestimentifera, and Pogonophora, (ii) in the Annelida phylum Hb_L disappeared, originating Hb and Chl, (iii) in Vestimentifera Hb_H evolved from Hb_L but the latter did not disappear, and (iv) Hb_H evolved separately from Hb and Chl. The last assumption is supported by the fact that vestimentiferans and pogonophorans are sister groups, independent from their systematic position. As a consequence, Hb_H could not be present in their common ancestor (an annelid) because otherwise we should admit its subsequent loss in Pogonophora, which contrasts with the starting hypothesis.

Our analyses performed on linkers of Hb, Chl, and Hb_H show that the genes coding for these polypeptides were already present in the common ancestor of Annelida, Vestimentifera, and Pogonophora. We therefore suggest an alternative scenario for the evolution of this group of respiratory pigments. Our most parsimonious hypothesis is based only on the following three assumptions (Fig. 7*B*). (i) The common ancestor of Annelida, Vestimentifera, and Pogonophora had in its genome the whole set of genes coding for globin and linker chains. As a consequence Hb, Chl, and Hb_H did not evolve independently. (ii) The common ancestor of Vestimentifera and Pogonophora evolved the Hb_L from Hb_H, starting from the set of genes coding for the latter. (iii) Pogonophora lost Hb_H during evolution.

The data deduced from our phylogenetic analyses fit nicely in this new hypothesis, which requires fewer *ad hoc* assumptions to explain the origin of this group of proteins. The new evolutionary model that we propose can be applied also to the Hbs of leeches, which were not considered in our study because the available sequences are not sufficiently complete. However, the data based on the amino termini of some leech globins clearly show that they are homologous to other extracellular Hbs (21).

Acknowledgments—We thank Dr. Lucio Cariello, Director of the Stazione Zoologica "A. Dhorn" (Napoli, Italy), for providing the specimens of *S. spallanzanii*.

REFERENCES

1. Vinogradov, S. N. (1985) *Comp. Biochem. Physiol. B* **82,** 1–15
2. Lamy, J. N., Green, B. N., Toulmond, A., Wall, J. S., Weber, R. E., and Vinogradov, S. N. (1996) *Chem. Rev.* **96,** 3113–3124
3. Gotoh, T., Shishikura, F., Snow, J. W., Ereifeji, J. W., Vinogradov, S., and Walz, D. A. (1987) *Biochem. J.* **241,** 441–445
4. Suzuki, T., and Riggs, A. F. (1993) *J. Biol. Chem.* **268,** 13548–13555
5. Yuasa, H. J., Green, B. N., Takagi, T., Suzuki, N., Vinogradov, S., and Suzuki, T. (1996) *Biochim. Biophys. Acta* **1296,** 235–244
6. Pallavicini, A., Negrisolo, E., Barbato, R., Dewilde, S., Ghiretti-Magaldi, A., Moens, L., and Lanfranchi, G. (2001) *J. Biol. Chem.* **276,** 26384–26390
7. Kapp, O. H., Moens, L., Vanfleteren, J., Trotman, C., Suzuki, T., and Vinogradov, S. (1995) *Protein Sci.* **4,** 2179–2190
8. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
9. Li, W. H. (1997) *Molecular Evolution*, Sinauer Associates, Inc., Sunderland, MA
10. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425
11. Conway Morris, S. (1993) *Nature* **361,** 219–225
12. Swofford, D. L. (1993) *Phylogenetic Analysis Using Parsimony*, Center for Biodiversity, Illinois Natural History Survey, Champaign, IL
13. Kluge, A. G., and Faris, J. F. (1969) *Syst. Zool.* **18,** 1–32
14. Faris, J. F. (1989) *Cladistics* **5,** 417–419
15. Madison, W. P., and Madison, D. R. (1992) *MacClade: Analysis of Phylogeny and Character Evolution*, Version 3, Sinauer Associates, Inc., Sunderland, MA
16. Van de Peer, Y., and De Wachter, R. (1994) *Comput. Appl. Biosci.* **3,** 227–230
17. Felsenstein, J. (1985) *Evolution* **39,** 783–791
18. Martin, P. D., Kuchumov, A. R., Green, B. N., Oliver, R. W. A., Braswell, E. H., Wall, J. S., and Vinogradov, S. N. (1996) *J. Mol. Biol.* **255,** 154–169
19. Suzuki, T., Takagi, T., and Ohta, S. (1990) *Biochem. J.* **266,** 221–225
20. Zal, F., Suzuki, T., Kawasaki, Y., Childess, J. J., Lallier, F. H., and Tulmond, A. (1997) *Proteins* **29,** 562–574
21. Shishikura, F., Ochiai, T., and Yamanaka, I. (1997) *Zool. Sci.* **14,** 923–930
22. Gotoh, T., Suzuki, T., and Takagi, T. (1991) in *Structure and Function of Invertebrate Oxygen Carriers* (Vinogradov S. N., and Kapp O. H., eds) pp. 279–283, Springer-Verlag New York Inc., New York
23. Takagi, T., Iwaasa, H., Ohta, S., and Suzuki, T. (1991) in *Structure and Function of Invertebrate Oxygen Carriers* (Vinogradov S. N., and Kapp O. H., eds) pp. 245–249, Springer-Verlag New York Inc., New York
24. Shishikura, F., and Nakamura, M. (1996) *Zool. Sci.* **13,** 849–856
25. Shishikura, F. (1996) *Zool. Sci.* **13,** 551–558

Downloaded from http://www.jbc.org/ by guest on July 25, 2018

The Evolution of Extracellular Hemoglobins of Annelids, Vestimentiferans, and Pogonophorans

Enrico Negrisolo, Alberto Pallavicini, Roberto Barbato, Sylvia Dewilde, Anna Ghiretti-Magaldi, Luc Moens and Gerolamo Lanfranchi

J. Biol. Chem. 2001, 276:26391-26397.
doi: 10.1074/jbc.M100557200 originally published online April 6, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M100557200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 20 references, 4 of which can be accessed free at
http://www.jbc.org/content/276/28/26391.full.html#ref-list-1

Downloaded from http://www.jbc.org/ by guest on July 25, 2018