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

# Dipartimento di Biologia and CRIBI Biotechnology Center, Università di Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy.
* Dipartimento di Scienze e Tecnologie Avanzate, Università del Piemonte Orientale Amedeo Avogadro, Corso Borsalino 54 I-15100 Alessandria, Italy
& Dipartimento di Biologia and CNR Center for the study of Metalloproteins, Università di Padova, via Ugo Bassi 58/B, 35131 Padova, Italy
§ Department of Biochemistry, University of Antwerp UIA, Antwerp, Belgium

! Corresponding author: Gerolamo Lanfranchi, Dipartimento di Biologia and CRIBI Biotechnology Center, Università di Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy
Tel: +390498276221; Fax: +390498276280; Email: lanfra@cribi.unipd.it

Running title: Evolution of annelid hemoglobins
Summary

The evolution of extracellular hemoglobin of Annelids, Vestimentiferans and Pogonophorans was investigated 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 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.
**Introduction**

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 an 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\(^1\) and Chls 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 Mw (~ 3000 kDa) to Hbs and Chls of Annelida, and is also composed of globin and linker chains. In contrast the Hb\(_L\) has a Mw of ~ 400 kDa and contains globin chains only. Pogonophora contain a 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 and collaborators divided the globin polypeptides in two groups with a common origin (3). Successively Suzuki and Riggs demonstrated that linker chains share similarities with the low-density lipoprotein receptor and that they cannot be related to the globins due to the high divergence of their primary structures (4). More recently Yuasa and co-workers have proposed a general evolutionary model for the Hbs, Hb\(_H\) and Hb\(_L\) (5). 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: 1) the placement of Chl in this...
Evolution of annelid hemoglobins

scenario and 2) the evolutionary pathway followed by Hbs, Chls and Hb\textsubscript{H} with respect to Hb\textsubscript{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\textsubscript{H}s and Hb\textsubscript{L}s. 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.
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) whereas 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 (http://www2.igh.cnrs.fr/bin/align-guess.cgi). Similar comparisons were carried out on the linker sequences using the ClustalW program (8).

Phylogenetic analyses

Sequence alignments were used to build phylogenetic trees were using the maximum parsimony (9) and the 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. They are: (i) the globins of the mollusc *Scapharca inaequivalvis*; (ii) the globin of the echiuran *Urechis caupo*; (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).
Evolution of annelid hemoglobins

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 exhaustive approach (12). All characters were equally weighted and those uninformative were excluded from the analysis. After a series of trials, we set the PAUP options as follows, because they performed more efficiently in finding the shortest trees. A) Optimization of the characters: ACCTRAN; B) Heuristic search options: B1) keep minimal trees only; B2) collapse zero-length branches; B3) starting tree sources: get by stepwise addition; B4) swap on: minimal trees only; B5) addition sequence: random; B6) replications: 50; B7) swapping algorithm: TBR tree bisection-reconnection; B8) MULPARS on; B9) 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 equally weighted. The PAUP setting was as follows. A) Optimization of the characters: ACCTRAN on; B) Exhaustive search options: B1) keep minimal trees only; B2) collapse zero-length branches on.

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 Index (C.I.) and the Retention Index (R.I.) (13, 14). These indexes were calculated as implemented in PAUP (12). Values of C.I. and R.I. > 0.5 indicate that convergent/parallel evolution does not affect strongly the phylogentic reconstruction and that the obtained topologies of trees are reliable. Molecular synapomorphies in the cladistic analysis were detected using the program MacClade version 3 that allows to track the evolution of each character (15).
Evolution of annelid hemoglobins

**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 1.3b (16). The NJ trees were created applying the following settings. A) Distance calculation: (*Kimura 83*); B) Alignment positions: *all*; C) Insertions & 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.
Evolution of annelid hemoglobins

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 in the Materials and Methods were added as outgroups. 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 BC corner, is not universal to all the globins. In two *Sabella* globins (Glb1 and Glb2) the small residues Ala and Cys are found at position B6 and E8. In these sites the majority of annelid globins have a Gly residue, with the exception of globin I of *Tylorrhynchus* where 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 position 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 databases (Fig. 2). The main feature that appear to be conserved is a
cysteine-rich segment \([(\text{Cys-}X_6)\text{-Cys-}X_5\text{-Cys-}X_10\text{-Cys})]\) which is typical for all the linker chains sequenced so far and has been related to the LDL-receptor motif (4). Considering the current global alignment (Fig. 2), the cysteine-rich segment can be written more precisely as follows \([(\text{Cys-}X_5\text{-}X_5\text{-}\text{Cys-Asp-}X_3\text{-Asp-Cys-X_4\text{-Asp-Glu-X_4-Cys}})]\).

**Pairwise comparison**

The analysis of the pairwise comparison of the globin cDNAs (Fig. 3) shows a higher percentage of identity in the ORF 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: 1) *Tylorrhynchus* L2 and *Neanthes* L2 markedly differ from the general structure of other linkers. In fact they share 74.58\% of identical amino acids, while the mean for the whole data set is 25.79\%. The latter value decreases to 23.35\%, if the comparison of *Tylorrhynchus* L2 vs. *Neanthes* L2 is excluded from the computation. 2) Roughly 40\% of the aminoacids are specific for each chain (mean of percentage of dissimilar amino acids = 39.45\%). 3) Identical aminoacids represent one forth 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**

*a) Globin phylogenetic analysis*

We have applied both cladistic and distance-based methods to study the molecular evolution of annelids globins. The cladistic analysis results in eighteen 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 in 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 appears to be controversial. In fact in the cladogram it is placed into the A group,
according to the classification of Gotoh et al. (3), while 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 1 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.

b) 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 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 relate the linkers with the LDL receptor (4), no other significant similarity is
know between linker chains and other proteins. Nevertheless, *Tylorrhynchus* L2 and *Neanthes*
L1 were used to root the trees, because they are much more similar between them 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. Conversely, the NJ tree supports the
two groups *Tylorrhynchus* L1 + *Lumbricus* L1 and *Sabella* L1 + *Sabella* L3. The first is also
Evolution of annelid hemoglobins

corroborated by the bootstrap value (72%), while 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 previous results of Gotoh et al. (3). These authors proposed the division of the globin chains in two main groups A and B, each divided further in the two subgroups A1/A2 and B1/B2. This classification was successively applied to HbHs and HbLs (19, 20). Recently, a correction of this nomenclature has been proposed which postulates to invert 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 appear 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 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, HbHs and HbLs 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
Evolution of annelid hemoglobins

members of the class Opisthochaeta within the phylum Annelida (20). The critical point is: what was the scenario in which Chls, Hbs, HbHs and HbLs evolved?

Yuasa et al. (5) previously suggested that Hb, Chl and HbH evolved from HbL, which is made only of globin chains, adding linker chains to form the final hexagonal bilayered structure. If we accept this hypothesis we must make four assumptions (Fig. 7, A): (A1) HbL was present in the common ancestor of Annelida, Vestimentifera and Pogonophora; (B1) in the Annelida phylum HbL disappeared, originating Hb and Chl; (C1) in Vestimentifera HbH evolved from HbL, the latter however did not disappear; (D1) HbH evolved separately from Hb and Chl. The last assumption is supported by the fact that Vestimentiferans and Pogonophorans are sister groups, independently from their systematic position. As a consequence, HbH could not be present in their common ancestor (an annelid) because otherwise we should admit its subsequent loss in Pogonophora, in contrast with the starting hypothesis.

Our analyses performed on linkers of Hb, Chl and HbH 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): (A2) 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 HbH did not evolve independently; (B2) the common ancestor of Vestimentifera and Pogonophora evolved the HbL from HbH, starting from the set of genes coding for the latter; (C2) Pogonophora lost HbH, during their evolution.

The data deduced from our phylogenetic analyses fit nicely in this new hypothesis that 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
Evolution of annelid hemoglobins

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).
Acknowledgements

Authors wish to thank Dr. Lucio Cariello Director of the Stazione Zoologica “A.Dhorn” (Napoli, Italy) for providing the specimens of Sabella spallanzanii. S.D. is a post-doctoral fellow of the Fund for Scientific Research-Flanders (FWO). This work was partly supported by the Italian Ministry of University and of Scientific and Technological Research (MURST grant - Cofinanziamento Prot. 9805192993-002 to A.G-M.).
References


Evolution of annelid hemoglobins


Evolution of annelid hemoglobins

Footnotes

+ E. Negrisolo and A. Pallavicini equally contributed to the work presented in this paper.

1 Abbreviations used are:

3'UTR, 3' untranslated region; C.I., consistency index; Chl, chlorocruorin; Hb, extracellular hemoglobin; HbH, heavy hemoglobin; HbL, light hemoglobin; LDL, low density lipoprotein; Mw, molecular weight; NJ, neighbor-joining; ORF, open reading frame; R.I., retention index; SC, strict consensus.

2 Glossary of the 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 which 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 where all conflicting branching patterns are collapsed into multifurcations. Synapomorphy, a derived state of a character that is shared by all the taxa.
Evolution of annelid hemoglobins

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.
Figure 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). Legend: 1) Scapharca I (P02213), Scapharca inaequalvis; 2) Scapharca IIA (P14821), Scapharca inaequalvis; 3) Urechis F1 (P06148), Urechis caupo; 4) Glycera MH4 (P15447), Glycera dibranchiata; 5) Glycera mmc (P02216), Glycera dibranchiata; 6) Glycera I (P23216), Glycera dibranchiata; 7) Glycera II (P21659), Glycera dibranchiata; 8) Glycera III (P21660), Glycera dibranchiata; 9) Oligobrachia b (5), Oligobrachia mashikoi; 10) Tylorrhynchus I (P02219), Tylorrhynchus heterochaetus; 11) Tylorrhynchus IIA (P09966), Tylorrhynchus heterochaetus; 12) Lamellibrachia BI (23), Lamellibrachia sp.; 13) Oligobrachia a5 (5), Oligobrachia mashikoi; 14) Lamellibrachia AIII (P15469), Lamellibrachia sp.; 15) Riftia b (P80592), Riftia pachyptila; 16) Pontodrilus I (24), Pontodrilus matsushimensis; 17) Pheretima sie. I (P11740), Pheretima sieboldi; 18) Pheretima com. I (24), Pheretima comunissima; 19) Pheretima hil. I (25), Pheretima hilgendorfi; 20) Lumbricus d1 (U55073), Lumbricus terrestris; 21) Lumbricus d2 (U55074), Lumbricus terrestris; 22) Tubifex I (P18202), Tubifex tubifex; 23) Sabella Glb3 (AJ131285), Sabella spallanzanii; 24) Lumbricus II (P02218), Lumbricus terrestris; 25) Lamellibrachia BIV (23), Lamellibrachia sp.; 26) Lumbricus III (P11069), Lumbricus terrestris; 27) Tylorrhynchus IIB (P13578), Tylorrhynchus heterochaetus; 28) Tylorrhynchus IIC (P02220), Tylorrhynchus heterochaetus; 29) Lumbricus IV (P13579), Lumbricus terrestris; 30) Oligobrachia c (5), Oligobrachia mashikoi; 31) Lamellibrachia BII (23), Lamellibrachia sp.; 32) Sabellastarte E (D58418), Sabellastarte indica; 33) Sabella Glb1 (AJ131283), Sabella spallanzanii; 34) Sabella Glb2 (AJ131284), Sabella spallanzanii. The accession numbers for the Swiss-Prot, NCBI and EMBL Data Banks are reported in brackets.
The amino acids common among all the globins of Annelida, Pogonophora and Vestimentifera are indicated by an asterisk. \( \equiv \) = invariant amino acids; \( \equiv \equiv \) = amino acids common to the majority of sequences; \( \equiv \equiv \equiv \) = synapomorphic amino acids; \( \equiv \equiv \equiv \equiv \) = autoapomorphic amino acids.

**Figure 2.** Alignment of linker chains.

Amino acid sequences of the following species were used: 1) *Tylorrhynchus* L2 (P18208), *Tylorrhynchus heterochaetus*; 2) *Neanthes* L2 (D58413), *Neanthes diversicolor*; 3) *Sabella* L1 (AJ131900), *Sabella spallanzanii*; 4) *Lumbricus* L1 (A46587), *Lumbricus terrestris*; 5) *Tylorrhynchus* L1 (P18207), *Tylorrhynchus heterochaetus*; 6) *Lamellibrachia* LAV1 (P16222), *Lamellibrachia* sp.; 7) *Sabella* L3 (AJ131286), *Sabella spallanzanii*. The accession numbers in Swiss-Prot, NCBI and EMBL Data Banks are reported in brackets. \( \equiv \equiv \) = invariant amino acids; \( \equiv \equiv \equiv \) = amino acids common to the majority of sequences.

**Figure 3.** Pairwise comparison of ORF and 3'UTR regions of globin cDNAs.

Black bars = globin ORF; white bars = globin 3'UTR region. 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.

**Figure 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. 1) TY2 = *Tylorrhynchus* L2 linker; 2) NEA = *Neanthes* L2 linker; 3) SA1 = *Sabella* L1 linker; 4) LUM = *Lumbricus* L1 linker; 5) TY1 = *Tylorrhynchus* L1 linker; 6) LAM = *Lamellibrachia* LAV1 linker; 7) SA3 = *Sabella* L3 linker.
**Figure 5.** Phylogenetic analysis of globin polypeptides.

A). The SC tree derived from merging eighteen most parsimonious cladograms. The statistic analyses of the most parsimonious cladograms gave the following values: Length = 1689 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 percentage after 500 replicates; only values ≥ 50% are reported. Lower case letters located close to the nodes refer to monophyletic clades that are supported by the molecular synapomorphies reported in Table 1. The standard abbreviations of globins are shown in brackets according to the nomenclature previously suggested by Gotoh *et al.* (22).

**Figure 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; C.I. = 0.862; R.I. = 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 in percentage after 500 replicates; only values ≥ 50% are reported.

**Figure 7.** Evolutionary models of extracellular hemoglobins.

In this figure are presented the two alternative hypotheses for the evolution of extracellular hemoglobins. Letters in brackets correspond to the different assumptions that support the two different models. See text for discussion.
Table 1. 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.

<table>
<thead>
<tr>
<th>i</th>
<th>Clade</th>
<th>BT</th>
<th>Molecular Synapomorphies</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td><em>Tylorrhynchus</em> IIA + <em>Tylorrhynchus</em> I</td>
<td></td>
<td>Gln (20), Val (29), Asn (52).</td>
</tr>
<tr>
<td>b</td>
<td><em>Lamellibrachia</em> AIII + <em>Riftia</em> b</td>
<td>100%</td>
<td>Tyr (8), Phe (42), His (46), Val (47), Thr (74), Thr (82), Thr (103), Val (129), His (131), Met (135), Gln (168).</td>
</tr>
<tr>
<td>c</td>
<td><em>Pontodrilus</em> I + (<em>Pheretima</em> sie. I + (<em>Pheretima</em> com. I + <em>Pheretima</em> hig. I))</td>
<td>95%</td>
<td>Leu (54), Glu (97), Pro (169).</td>
</tr>
<tr>
<td>d</td>
<td><em>Pheretima</em> sie. I + (<em>Pheretima</em> com. I + <em>Pheretima</em> hig. I)</td>
<td>96%</td>
<td>His (42), Phe (43).</td>
</tr>
<tr>
<td>e</td>
<td>Hbs+Chls+Hbds+Hbds</td>
<td>98%</td>
<td>Cys (10), Arg (83), Cys (159)</td>
</tr>
<tr>
<td>f</td>
<td><em>Pheretima</em> com. I + <em>Pheretima</em> hig. I</td>
<td>68%</td>
<td>Glu (29), His (136), Phe (156).</td>
</tr>
<tr>
<td>g</td>
<td><em>Pontodrilus</em> I to <em>Lumbricus</em> d2;</td>
<td></td>
<td>Asp (139), Thr (149)</td>
</tr>
<tr>
<td>h</td>
<td><em>Lumbricus</em> d1 + <em>Lumbricus</em> d2</td>
<td>98%</td>
<td>Pro (57), Thr (97), Met (100)</td>
</tr>
<tr>
<td>i</td>
<td><em>Tylorrhynchus</em> IIB to <em>Sabella</em> Glb1</td>
<td></td>
<td>Ser (35).</td>
</tr>
<tr>
<td>j</td>
<td><em>Sabellastarte</em> E + (<em>Sabella</em> Glb2 + <em>Sabella</em> Glb1)</td>
<td>100%</td>
<td>Met (12), Ala (21), Glu (23), Lys (50), Ala (95), Phe (96), Met (121), Leu (126), Ala (140), Pro (144), Ala (162), Met (163).</td>
</tr>
<tr>
<td>k</td>
<td><em>Sabella</em> Glb2 + <em>Sabella</em> Glb1</td>
<td>98%</td>
<td>Asn (110), Ser (170).</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Sabella L</strong></td>
<td><strong>Tylorrhynchus L</strong></td>
<td><strong>Neanthes L</strong></td>
<td><strong>Sabella L</strong></td>
</tr>
<tr>
<td><strong>L3</strong></td>
<td><strong>L2</strong></td>
<td><strong>L1</strong></td>
<td><strong>L1</strong></td>
</tr>
<tr>
<td><strong>L3</strong></td>
<td><strong>L2</strong></td>
<td><strong>L1</strong></td>
<td><strong>L1</strong></td>
</tr>
<tr>
<td><strong>Tylorrhynchus L</strong></td>
<td><strong>Neanthes L</strong></td>
<td><strong>Sabella L</strong></td>
<td><strong>Lumbricus L</strong></td>
</tr>
<tr>
<td><strong>L2</strong></td>
<td><strong>L2</strong></td>
<td><strong>L3</strong></td>
<td><strong>L1</strong></td>
</tr>
<tr>
<td><strong>L2</strong></td>
<td><strong>L3</strong></td>
<td><strong>L1</strong></td>
<td><strong>L1</strong></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td><strong>2</strong></td>
<td><strong>3</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

**Figure 2**
Figure 4
Figure 7
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. 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 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2001/04/06/jbc.M100557200.citation.full.html#ref-list-1