CHAPTER THIRTY

Inferring Evolution of Fish Proteins: The Globin Case Study

Agnes Dettaï,* Guido di Prisco,* Guillaume Lecointre,* Elio Parisi,* *and* Cinzia Verde†

Contents

* UMR, Département Systématique et Evolution, Muséum National d'Histoire Naturelle, Paris, France
† Institute of Protein Biochemistry, CNR, Naples, Italy

Methods in Enzymology, Volume 436
ISSN 0076-6879, DOI: 10.1016/S0076-6879(08)36030-3

Abstract

Because hemoglobins (Hbs) of all animal species have the same heme group, differences in their properties, including oxygen affinity, electrophoretic mobility, and pH sensitivity, must result from the interaction of the prosthetic group with specific amino acid residues in the primary structure. For this reason, fish globins have been the object of extensive studies in the past few years, not only for their structural characteristics but also because they offer the possibility to investigate the evolutionary history of Hbs in marine and freshwater species living in a large variety of environmental conditions. For such a purpose, phylogenetic analysis of globin sequences can be combined with knowledge of the phylogenetic relationships between species. In addition, Type I functional-divergence analysis is aimed toward predicting the amino acid residues that are more likely responsible for biochemical diversification of different Hb families. These residues, mapped on the three-dimensional Hb structure, can provide insights into functional and structural divergence.

1. Introduction

The repertoire of known globin sequences has been steadily increasing in the past years, making possible a comparative study on hemoglobin (Hb) in a vast variety of species. The number of available fish globin sequences provides enough material to study evolutionary and functional aspects. Fish Hbs are particularly interesting because the respiratory function of fish differs from that of mammals. In fish, gills are in contact with a medium endowed with high oxygen tension and low carbon-dioxide tension; in contrast, in the alveoli of mammalian lungs, the carbon dioxide tension is higher and the oxygen tension is lower than in the atmosphere. Unlike most mammals, including humans, fish exhibit Hb multiplicity, which results from gene-related heterogeneity and gene duplication events. An important feature of vertebrate Hbs is the decreased oxygen affinity at low pH values (Riggs, 1988), known as the Bohr effect. In many teleost Hbs, the oxygen affinity is so markedly reduced at low pH that Hb cannot be fully saturated even when oxygen pressure is very high (Root effect). This effect (Brittain, 2005) plays an important physiological role in supplying oxygen to the swim bladder and choroid rete mirabile. Thus, Root effect Hbs can regulate both

buoyancy and retina vascularization (Wittenberg and Wittenberg, 1974). The Root effect apparently evolved 100 million years before the appearance of the choroid rete (Berenbrink *et al.*, 2005). Interestingly, weakening in the intensity of the Root effect has been noticed among the Antarctic notothenioids, although some species retain a strong effect (di Prisco *et al.*, 2007).

A common characteristic among fish is Hb multiplicity, usually interpreted as a sign of phylogenetic diversification and molecular adaptation. From the phylogenetic viewpoint, teleost Hbs have been classified as embryonic Hbs, Antarctic major adult Hbs, anodic adult Hbs, and cathodic adult Hbs (Maruyama *et al.*, 2004). Embryonic Hbs are typical of the growing embryo and of the juvenile stages. The group of Antarctic major adult Hbs (Hb 1) includes the globins of red-blooded notothenioids, the dominant fish group in Antarctica, but also some globins of temperate species. In Antarctic notothenioids, Hb 1 accounts for 95 to 99% of the total, while the juvenile/embryonic forms are normally present in trace amounts in the adults. Presumably, the two clusters of Antarctic major and embryonic Hbs were generated by gene-duplication events, which occurred independently for the α- and β-globin genes. According to Bargelloni *et al.* (1998), the duplication event that gave origin to the two groups of Antarctic globins involved a mechanism of positive selection (i.e., changes that improve the fitness of the species), characterized by higher rate of nonsynonymous (amino acid replacing) to synonymous (silent) substitutions. Cathodic Hbs are considered to play an important role in oxygen transport under hypoxic and acidotic conditions. The embryonic Hb group also includes the α-globin sequences of *Anarhichas minor* Hb 1, and the β-globin sequences of *A. minor* Hb 3, *Liparis tunicatus* Hb 1, and *Chelidonichthys kumu* Hb (Giordano *et al.*, 2006).

It has been suggested that Hb multiplicity is more frequently found in fish that must cope with variable temperatures, whereas the presence of a single dominant Hb is usually associated with stable temperature conditions. This may explain why high-Antarctic notothenioids have a single major Hb, while sub-Antarctic and temperate notothenioids, such as *Cottoperca gobio* and *Bovichtus diacanthus*, respectively, retained Hb multiplicity, presumably to cope with the small or large temperature changes in the respective habitats north of the polar front (di Prisco *et al.*, 2007).

Although a report in the literature (Sidell and O'Brien, 2006) does not support the ensuing hypothesis, we believe that the reduction in Hb content/multiplicity and erythrocyte number in the blood of high-Antarctic notothenioids counterbalances the potentially negative physiological effects (e.g., higher demand of energy needed for circulation) caused by the increase in blood viscosity produced by subzero seawater temperature.

As detailed in this chapter, the polar ecosystems offer numerous examples of the evolution of the relationships between globin structure and function among fish. The approaches herewith outlined are based on interplay of phylogeny with globin-structure features. To date, this multidisciplinary

approach has not been sufficiently exploited. New information will help us to understand the evolution of globins and to address the question whether different structures and functions are related to environmental conditions.

2. Hb Purification

Rigorous purification and primary-structure analyses are fundamental to properly study the evolutionary history of globin genes and proteins, and several methodologies have been developed over the years. In the field of fish Hbs, a detailed overview of hematology and biochemical methods is beyond the goal of this contribution; the reader is directed to Antonini *et al.* (1981) and Everse *et al.* (1994). The procedure for Hb purification given subsequently has been employed for the isolation and functional characterization of the three major Hbs of the Arctic gadid *Arctogadus glacialis* (Verde *et al.*, 2006).

2.1. Hemolysate preparation

Blood erythrocytes are packed by low-speed centrifugation for 5 min, washed three times with 1.7% NaCl in 1 m*M* Tris-HCl, pH 8.1, and lysed in 1 m*M* Tris-HCl, pH 8.1. Following centrifugation at 100,000×*g* for 60 min, the supernatant is dialyzed against Tris-HCl, pH 7.6. All manipulations are carried out at 0 to 4°. For functional studies, endogenous organophosphates are removed from the hemolysate by passage through a small column of Dowex AG 501 X8 (D) (e.g., stripping). No oxidation was spectrophotometrically detectable during the time needed for functional experiments. Hb solutions were stored in small aliquots at −80° until use.

2.2. Hb separation

A column of Mono Q HR5/5 (Pharmacia) is equilibrated with 10 m*M* Tris-HCl, pH 7.6 (buffer A). Hb 3 is eluted at 30% buffer B (250 m*M* Tris-HCl, pH 7.6, containing 250 m*M* NaCl) at a flow rate of 1.0 ml/min. Figure 30.1 (modified from the work of Verde *et al.*, 2006) shows separation of the three Hbs in the hemolysate of *A. glacialis* carried out by fast protein liquid chromatography (FPLC).

3. Elucidation of Globin Primary Structure

3.1. Globin separation

The hemolysate or solutions of purified Hbs (35 mg/ml) are incubated at room temperature for 10 min in a denaturing solution containing 5% β-mercaptoethanol and 1% trifluoroacetic acid (TFA). High-performance

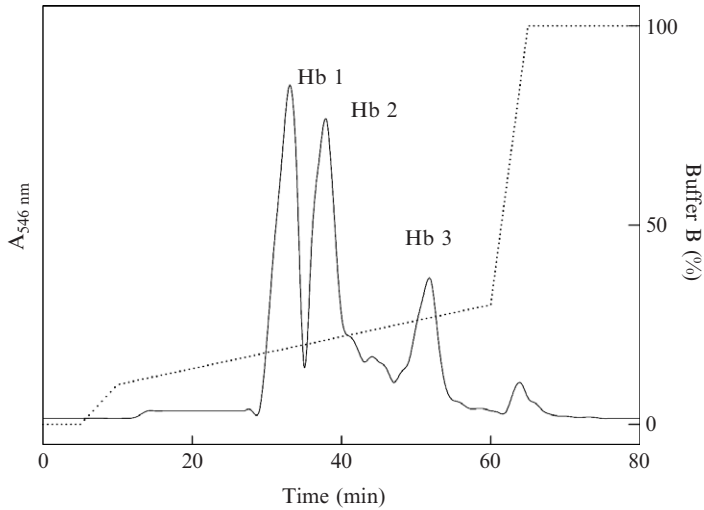

Figure 30.1 Ion-exchange chromatography of the hemolysate of *A. glacialis* (modified from Verde *et al.*, 2006).

liquid chromatography (HPLC) is carried out at room temperature (0.5 mg for each run) on a C_4 reverse-phase column, equilibrated with 45% acetonitrile, 0.3% TFA (solvent A), and 90% acetonitrile, 0.1% TFA (solvent B). The column is eluted at a flow rate of 1.0 ml/min and the eluate is monitored at 280 and 546 nm (Fig. 30.2). The procedure for globin purification described has been employed for the isolation of the four globins from the Arctic gadid *Gadus morhua* (Verde *et al.*, 2006). Fractions containing the eluted globin chains are pooled and lyophilized.

3.2. Modification of α and β chains of fish hemoglobins

Cysteyl residues are alkylated with 4-vinyl-pyridine (Friedman *et al.*, 1970). Lyophilized globins are dissolved at a concentration of 10 mg/ml in 0.5 *M* Tris-HCl, pH 7.8, containing 2 m*M* EDTA and 6 *M* guanidine hydrochloride under nitrogen. A 5-fold excess of dithiothreitol is added under nitrogen; the solution is kept at 37° for 2 h; incubation with 20-fold excess of 4-vinyl-pyridine is then carried out for 45 min at room temperature in the dark under nitrogen. The reaction is stopped by addition of 2.5-fold molar excess of dithiothreitol. Excess reagents are removed by reverse-phase HPLC. The eluted *S*-pyridylethylated α and β globins are lyophilized.

3.3. Protein cleavage

Tryptic digestion is carried out on pyridylethylated chains (100 nmol/ml), dissolved in 50 m*M* Tris-HCl, pH 8.0 (Tamburrini *et al.*, 1992, 1996). Trypsin (1 mg/ml, dissolved in 1 m*M* HCl), is added at a ratio of 1:100

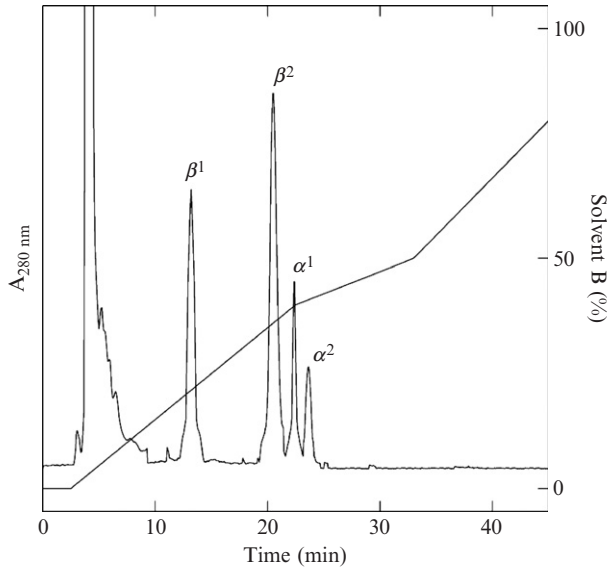

Figure 30.2 Reverse-phase HPLC of the hemolysate of *G. morhua* (Verde *et al.*, 2006), containing four globins. A C_4 Vydac column (4.6 × 250 mm) is equilibrated with solvent A. Elution is performed with a gradient of 90% solvent B in solvent A.

(by weight); after 3-h incubation at 37°, another aliquot of enzyme is added, at a final concentration of 1:50 (by weight). After 6 h, the reaction is stopped by lyophilization. The hydrolysate is suspended in 0.1% TFA and clarified by centrifugation. Separation of tryptic peptides (Tamburrini *et al.*, 1992) is carried out by HPLC on a μ-Bondapack C_{18} reverse-phase column equilibrated with 0.1% TFA in water (solvent A) and 0.08% TFA in 99.92% acetonitrile (solvent B). Elution is carried out at a flow rate of 1 ml/min, and the eluate is monitored by measuring the absorbance at 220 and 280 nm. Up to 20 nmol of peptide mixture can be loaded in each preparative run. Peptides are manually collected and dried in a Savant speed concentrator. Cleavage of Asp-Pro bonds is performed on intact globins in 70% (v/v) formic acid, for 24 h at 42° (Landon, 1977).

3.4. Deacylation of α-chain N terminus

Partial deacylation (approximately 20%) of the blocked N terminus of the α chains occurs spontaneously during Asp-Pro cleavage, due to the acidic reaction. Additional deacylation (50%) is obtained by incubating the N-terminal tryptic peptide with 30% TFA for 2 h at 55°. The N-terminus-blocking group is identified by matrix-assisted laser-desorption

ionization-time-of-flight (MALDI-TOF) mass spectrometry on a PerSeptive Biosystems Voyager-DE Biospectrometry Workstation (Verde *et al.*, 2006).

4. Globin Characterization

4.1. Amino acid sequencing

Sequencing is performed on *S*-pyridylethylated globins, on fragments generated by Asp-Pro cleavage, and on HPLC-purified tryptic peptides with an automatic sequencer equipped with online detection of phenylthiohydantoin amino acids. Sequencing of Asp-Pro-cleaved globins is performed after treatment with *o*-phthaldehyde (OPA) (Brauer *et al.*, 1984) in order to block the non-Pro N terminus and reduce the background.

4.2. Cloning and sequencing of globin cDNAs

The primary structure of globins can also be deduced from the sequence of cDNA. This procedure offers some advantages because it is less time consuming than direct protein sequencing. In addition, the knowledge of the nucleotide sequence can be useful for evolutionary studies, especially to establish instances of positive selection, which can be inferred by measuring the ratio of nonsynonymous over synonymous substitutions. However, the knowledge of the sequence of the N-terminal peptide can be useful for the preparation of specific templates to be used for PCR. The protocol given below has been employed for nucleotide sequencing of α and β globins of *A. glacialis* (Verde *et al.*, 2006). Total spleen RNA is isolated using TRI Reagent (Sigma-Aldrich), as described in Chomczynski and Sacchi (1987). First-strand cDNA synthesis is performed according to the manufacturer's instructions (Promega), using an oligo(dT)-adaptor primer. The α- and β-globin cDNAs are amplified by PCR using oligonucleotides designed on the N-terminal regions as direct primers and the adaptor primer as the reverse primer. Amplifications are performed with 2.5 units of Taq DNA polymerase, 5 pmol each of the above primers, and 0.20 m*M* dNTPs buffered with 670 m*M* Tris-HCl, pH 8.8, 160 m*M* ammonium sulfate, 0.1% Tween 20, and 1.5 m*M* $MgCl_2$. The PCR program consists of 30 cycles of 1 min at 94°, 1 min at temperatures between 42 and 54°, 1 min at 72°, and ending with a single cycle of 10 min at 72°. The cloning of the N-terminal regions is obtained by 5′ RACE (rapid amplification of cDNA ends), using the Marathon cDNA Amplification Kit (BD Biosciences) and two internal primers. Amplified cDNA is purified and ligated in the pDrive vector (Qiagen). *Escherichia coli* cells (strain DH5{I}α{/I}) are transformed with the ligation mixtures. Standard molecular-biology techniques (Sambrook *et al.*, 1989) are used in the isolation, restriction, and sequence

analysis of plasmid DNA. When required, automatic sequencing is performed on both strands of the cloned cDNA fragments.

5. Sequence Analysis

As we will detail in subsequent sections, information about species history is highly valuable, and indeed essential in most cases for the interpretation of any observed change along the sequences from gene families and for the inference of duplications and losses of copies from these families. This requires gathering and analyzing relevant sequence data, as well as integrating knowledge of the phylogeny deriving from external sources.

5.1. Homologous and paralogous copies

In any comparative or phylogenetic analysis of sequences, the first step is to find the sequences relevant for comparison with regard to the question addressed. For phylogenetic purposes, it is necessary only to integrate sequences with a common origin. There are two kinds of such sequences. In one case, the divergence of an ancestral lineage in two groups leads to one copy in each of the daughter lineages (orthologous copies; see also homology *sensu strictu* [Fitch, 1970]). In the other case, a duplication occurs within the genome of a single organism, whose descendants will thereafter possess two copies (paralogy [Fitch, 1970]). These two copies may diverge more or less independently by the usual process of accumulation of mutations, but the presence of redundancy (two copies are present but only one is necessary for the correct functioning of the organism) allows for several different outcomes and constitutes one of the main sources for genetic innovation in the genomes (Hurles, 2004; Lynch and Connery, 2000). Very often, one of the copies will simply become inactive (pseudogene). Sometimes, it can acquire a new function, either completely new (Long *et al.*, 2003; Prince and Pickett, 2002) or linked to the original one (Ohno, 1970), sometimes even leading to subfunctionalization of the original function between the two divergent copies (Force *et al.*, 1999). The study of the evolution of paralogues is a complex but highly productive approach that can be improved by introducing phylogenetic analysis and other information in addition to the analytical comparative methods (for a review, see Mathews, 2005). Phylogenetic analyses, as all comparative methods in molecular studies, need an alignment of the relevant sequences.

5.2. Sequence search in the databases

Sequences can either be obtained through benchwork, as described previously, or retrieved from sequence databases such as GenBank or EMBL through various portals (http://www.ncbi.nlm.nih.gov/, http://srs.ebi.ac.uk/), and

from the complete genomes (http://www.ensembl.org/). Query by gene name and taxon name is a good way to start but will generally not recover all existing sequences for a given gene. Some of the sequences might not yet be annotated (the name of the gene is not yet associated with this sequence in the database), and others might be filed under a different name. To recover a maximal number of sequences, it is wise to also search with tools like BLAST (http://www.ncbi.nlm.nih.gov/BLAST/; http://www.ensembl.org/Multi/blastview). These rely on comparison among the sequences, not on the annotation. Adjusting the BLAST parameters to the type of search is very important, because too-stringent parameters will lead the program to ignore more divergent sequences, including paralogues of interest. In contrast, too widely defined search parameters will yield many irrelevant sequences and need much subsequent sorting. Detailed help can be found at http://www.ncbi.nlm.nih.gov/blast/producttable.shtml.

5.3. Sequence alignment

Sequence alignment is the way to arrange amino acid and nucleotide sequences in order to exhibit correspondances among regions with similarities. Therefore, obtaining a reliable alignment of the sequences is essential. In a first step, it is assumed that the similarity ensues from common ancestry of the residues present at a given position in the aligned sequences (homology hypotheses within the sequence). The phylogenetic tree will allow sorting of which of the similar residues at corresponding positions are really inherited from a common ancestor and which are similar due to convergences. The alignment process is often complicated by the fact that sequences of different species or different members of the gene family of interest are of different lengths. Pairing must then be obtained by introducing dashes (insertions/deletions) at carefully chosen places in order to maximize similarities. Shared similarity can result from shared ancestry, from convergence (homoplasy) led by functional or structural constraints, or even from purely contingent reasons. This can also lead to errors. Errors in the alignment can lead to errors in the phylogenetic reconstructions and in the conclusions drawn about adaptive features.

There are multiple methods and computer programs to align sequences. For all of them, it is necessary to carefully consider the parameters (generally a gap penalty and a substitution matrix to calculate the penalty of all possible misassociations). Schematically, these are used to calculate the value of a given alignment over another, as changes in the parameters may cause large differences in the alignment; for a more developed explanation, see Page and Holmes (1998) or Nei and Kumar (2000). One very widely used tool is Clustal (Higgins and Sharp, 1988). It has several more recent and complete versions, for instance, ClustalX (Thompson *et al.*, 1997), some of which

have been integrated in other programs, free or commercially available (Bioedit [Hall, 1999]; GCG, Accelrys; etc). Clustal is available for use on many online servers (e.g., the EMBL-EBI server, http://www.ebi.ac.uk/clustalw/).

More efficient programs have been developed over the years, such as T-coffee (Notredame *et al.*, 2000) or 3DCoffee (O'Sullivan *et al.*, 2004), the latter even being able to integrate three-dimensional structures in the alignment estimations. For highly similar sequences, it is also possible to align by hand (for guidelines, see Barriel, 1994). For more distantly related globin sequences, the more effective, although slower, progressive method implemented in the software T-Coffee (Notredame *et al.*, 2000) can be used to produce more accurate alignments than ClustalW. Alignment of fish β globins with globins from an arthropode and a nematode worm (*Daphnia magna* and *Pseudoterranova decipiens*) obtained with T-Coffee is shown in Figure 30.3.

Nucleic acid sequences can be aligned as well as amino acid residues sequences, but the parameters and cost matrices must be adjusted accordingly. When aligning nucleic acid–coding sequences by hand, the reading frame must be considered and conserved, especially if the sequences will be used in a functional study. The corresponding amino acid residue alignment can be used as a guideline for manual editing if necessary. Some software allows for easy switching between amino acid–residue and nucleic acid alignments (e.g., Se-Al, http://evolve.zoo.ox.ac.uk/, or Bioedit [Hall, 1999]).

6. Phylogenetic Analysis

6.1. Overview of tree-building methods and topology interpretation

Phylogenetic analysis is performed on the matrix of homology hypotheses (sequence alignment). Two main groups of approaches exist for this process: a tree is either constructed from a distance matrix among sequences (distance methods) or selected within the space of possible trees according to an optimality criterion that is dependent on the method (parsimony and maximum-likelihood methods). Maximum-parsimony (MP) and maximum-likelihood (ML) methods allow for direct character and character-change mapping on the trees. Many books and review papers explore in depth the choice and the use of the various methods, including their pitfalls (Felsenstein, 2004; Nei and Kumar, 2000; Page and Holmes, 1998). Nonetheless, it is important to keep in mind that the distance trees are actually similarity trees, where the topology does not directly reflect common ancestry, but rather degrees of global similarity between sequences. These trees, called phenograms, are obtained through methods such as neighbor joining (NJ [Saitou and Nei, 1987]) or UPGMA. With

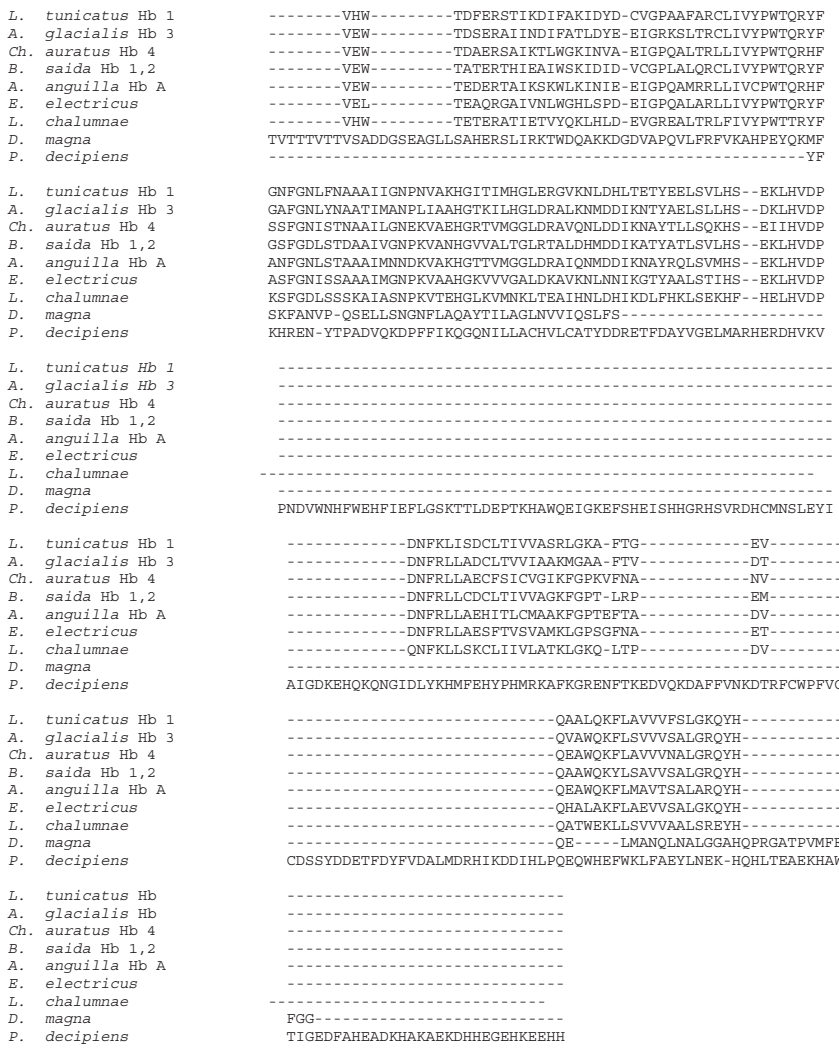

Figure 30.3 Alignment of sequences of fish β globins with two invertebrate globins. The alignment was created by the software T-COFFEE Version 1.41.

these, the tree is calculated from a distance matrix established by comparison of the sequences two by two. Phenetic trees can reflect phylogenetic topologies when the similarity in sequences is proportional to kinship (i.e., when sequences evolve at the same constant rate among lineages). Corrections can be introduced, when this is not the case, by the use of models (Nei and Kumar, 2000). Also, more effective distance methods have been proposed (e.g., Gascuel and Steel, 2006).

MP, ML, and Bayesian approaches (BA) also have their own pitfalls. As the tree-space explorations are done by a heuristic search, the approach does not explore all possible trees. Therefore, in some cases, the approach can remain stuck in a local optimum and never find the most parsimonious tree or the tree with the highest likelihood. This danger can be limited by making several analyses from different starting trees. Guidelines on how to optimize the search and the number of replicates are generally available in the documentation of the phylogenetic reconstruction programs, and more efficient search methods have been proposed. The parsimony ratchet method (Nixon, 1999a) is highly efficient for fast parsimony analyses and is implemented in several programs such as TNT (Goloboff *et al.*, 2000) and PAUPRat (Sikes and Lewis, 2001), one of the available ratchet-implementing programs for the widely used PAUP* (Swofford, 2001). For ML and BA, as well as for model-using distance methods, the choice of the most adapted model is crucial (Page and Holmes, 1998). The most widely used programs are ModelTest (Posada and Buckley, 2004; Posada and Crandall, 1998) and ProtTest (Abascal *et al.*, 2005). For the analyses themselves, fast programs like Mr. Bayes (Ronquist and Huelsenbeck, 2003) and PhyML (Guindon and Gascuel, 2003) are available and widely used. Even if the speed of analysis is still lower than that for a distance analysis, the time needed is generally quite reasonable (from a few hours to a few days, depending on the program), and it is worth using other methods in addition to distance. Bootstrap analyses must also be performed on a high number of replicates, as the variance in the results is high when only a few are performed.

The phenomenon named long-branch attraction (LBA [Hendy and Penny, 1989]) has been demonstrated when the sequences change at unequal rates among branches, even when unlimited data is available. Taxa with higher rates of mutation (i.e., long branches) tend to attract one another in the inferred tree. Schematically, taxa evolving at higher rates tend to have similar character states by convergence, and tree-building methods tend to group them together (Page and Holmes, 1998). As a result, longer branches of these taxa are often attracted to the outgroup, which frequently has the longest branches, as it is more distantly related to the others. The choice of an outgroup, as closely related to the group of interest as possible, is important also for this reason. The lower the number of branches in a tree, the sharper the problem; trees that include only four species are particularly sensitive to this artefact (Lecointre *et al.*, 1993; Philippe and Douzery, 1994). Some methods are more sensitive than others to this phenomenon, but models can be used to detect and partially correct it (see Lartillot *et al.*, 2007), although there is still widespread debate on the extent of the problem and sensitivity of the various methods. In general, whenever possible, adding sequences to the branches that are supposed to be

subject to LBA is also a good strategy ("breaking the long branches" [Graybeal, 1998; Hillis *et al.*, 2003]).

Additionally, in most cases, applying several methods to the data set is valuable, as discrepancies can provide information about which nodes are less supported (robustness to changes of method and model).

6.2. Approaching the species tree

A short review of the factors known to influence the inference of the species tree is of interest, as it shows where the reconstruction of a gene tree differs from it, and thereby the potential pitfalls specific to gene-tree building and interpretations. We will concentrate here on four main points of direct relevance also to gene-tree reconstruction.

6.2.1. Single copy rather than gene families

For phylogenetic reconstruction of species trees, it is desirable to choose marker(s) present as a single copy in the genome or at least to ensure that the various copies are clearly distinguishable from one another. When this is not the case, it can be difficult or impossible to distinguish between species-separation and gene-duplication events, and the interpretation of the tree as an approximation of the species tree is compromised (Cotton, 2005; Page and Holmes, 1998). This problem is especially relevant in teleost fish, as the teleost genome has undergone an ancient duplication, and many genes are present in multiple copies (Hurley *et al.*, 2007; Robinson-Rechavi *et al.*, 2001, Taylor *et al.*, 2001), with additional duplications and sometimes differential loss of copies in the various lineages (e.g., Hashiguchi and Nishida, 2006; Kuo *et al.*, 2005; Taylor *et al.*, 2003). But traces of other rounds of genome duplication have been revealed in numerous actinopterygian lineages, for instance some Acipenseriformes (Dingerkus and Howell, 1976) and some catfishes (Uyeno and Smith, 1972). A database of homologies and paralogies for genes present in single copy in Sarcopterygians but multiple copies in Actinopterygians is available at http://www.evolutionsbiologie.uni-konstanz.de/Wanda/index.htm (Van de Peer *et al.*, 2002).

6.2.2. Multiple markers

Several studies have also shown that using several diversified sequences as markers considerably improves the approximation of the species tree (for a review, see DeSalle, 2005). Concatenating the sequences of several markers and analyzing them simultaneously is also recommended (DeSalle, 2005; Kluge, 1989; Wiens, 1998), although performing separate analyses of each marker also allows for exploration of the possible marker-specific biases and differences in the history of the genes (de Queiroz, 1993; Dettaï and Lecointre, 2004, 2005; Maddison, 1997).

6.2.3. Markers with different rates of evolution

Different parts of the genome and genes evolve at different rates. To establish a phylogeny of closely related species, it is necessary to use sequences evolving rapidly, such as mitochondrial or noncoding sequences; otherwise an insufficient number of variable sites will be available to infer the relationships. In contrast, for phylogenies of distantly related species, the use of fast-evolving markers poses alignment and phylogenetic-signal erasure problems. The latter are due to the fact that there are only four possible character states in nucleic acid sequences, and moreover, some changes occur more frequently than others. As time goes by, multiple substitutions can occur at a single position, changing the initial shared character state and obscuring phylogenetic relationships. Although it is possible to correct for multiple substitutions, by the use of substitution models or by the elimination of the most variable positions (a theme covered at length in all molecular phylogenetics textbook), this remains one of the main sources of phylogenetic reconstruction error and uncertainty. Therefore, using more conserved markers for deeper divergences is still considered very important (Graybeal, 1994). Moreover, including multiple markers with different levels of variability and analyzing them simultaneously can enhance the resolution of the tree by providing signals for both deeper and shallower levels (Hillis, 1987).

6.2.4. Evaluating reliability

Finally, it is difficult to evaluate how far a given phylogenetic reconstruction can be trusted. Various indicators exist to estimate to what extent the data analyzed support a node in the tree (e.g., bootstrap values, jackknife, Bremer support [Bremer, 1994], and others). But they only measure robustness (i.e., how the signal within the data resists perturbations of the data set) or the contradiction within the data set. They cannot evaluate how well the tree inferred using the marker reflects the species tree and can even indicate high values for nodes due solely to marker biases such as high GC content (see Chang and Campbell, 2000). Thus, another argument for using several markers is the following: if all of them yield the same groups, regardless of possible marker-specific biases, the common signal should come from the shared ancestry (Miyamoto and Fitch, 1995). This allows for a more direct evaluation of how well the obtained tree reflects the species tree (Dettaï and Lecointre, 2004).

6.3. Some reference species trees for Actinopterygians

Although the phylogeny of Actinopterygians is still far from fully resolved, important progress has been made in recent years, especially with the new and fast-expanding wealth of molecular data. The multigene, taxon-rich

trees are generally congruent to a large extent with one another. We provide a few references herein, but it is only possible to cover a fraction of the existing publications. A good starting point for a summary of the morphological and some molecular data is Nelson (2006).

For gnathostomes, for instance, Janvier (1996), Stiassny *et al.* (1996), Kikugawa *et al.* (2004), and Takezaki *et al.* (2003, 2004) ought to be consulted. There is still a debate over the basal relationships among Actinopterygians, with part of the molecular data (e.g., Hurley *et al.*, 2007; Kikugawa *et al.*, 2004; Lê *et al.*, 1993) supporting the classical Neopterygii group, and another part (Inoue *et al.*, 2003; Venkatesh *et al.*, 2001) supporting an ancient fish clade, grouping chondrosteans, lepisosteids, and amiids. Within teleosts, a few references are Lê *et al.* (1993), Zaragueta-Bagils *et al.* (2002), Ishiguro *et al.* (2003), and Filleul and Lavoue (2001). Within Acanthomorpha, see Chen *et al.* (2003), Dettaï and Lecointre (2004, 2005), Miya *et al.* (2003, 2005), and Smith and Wheeler (2004, 2006).

6.4. Reconstruction of the history of a gene family

The reconstruction of the phylogeny of the gene of interest does not fulfill the four requirements listed earlier. In the case of a gene tree, the gene family of interest itself imposes the choice of the sequences to be included. It is generally not possible to select other markers more adapted to the time of divergence or to add more markers to counteract possible biases or too-short sequences containing not enough information, because it cannot be excluded that these other genes have a different history (see Figs. 30.4 to 30.6). The case of the tree obtained from the partial α globins in Near *et al.* (2006) is a good example of lack of resolution inevitable from the shortness of the sequences and the recent divergence of the icefishes. These resulted in a very low number of variable sites, sufficient to resolve only some parts of the tree. Nonetheless, it is not wise to include additional sequence length, as the analysis of the β globins shows that they had a different history from the α globins.

It is therefore important to be very careful when aligning and inferring the gene family tree. The tree can only be checked indirectly against other trees from other markers, and some errors might not be detectable in the way they would be if optimizing for species-tree recovery among several meticulously selected genes. The inclusion of both genes and pseudogenes in a single data set can be especially tricky and is a common occurrence in gene family trees. Pseudogenes generally evolve at higher rates; the clustering of all pseudogenes together may come from LBA and must be checked. While other sequences from the same organism cannot be integrated in the analysis, it is possible to avoid some pitfalls of phylogenetic reconstruction and clarify some parts of the tree by sampling a large number of organisms, representative of as many clades as possible of known species trees.

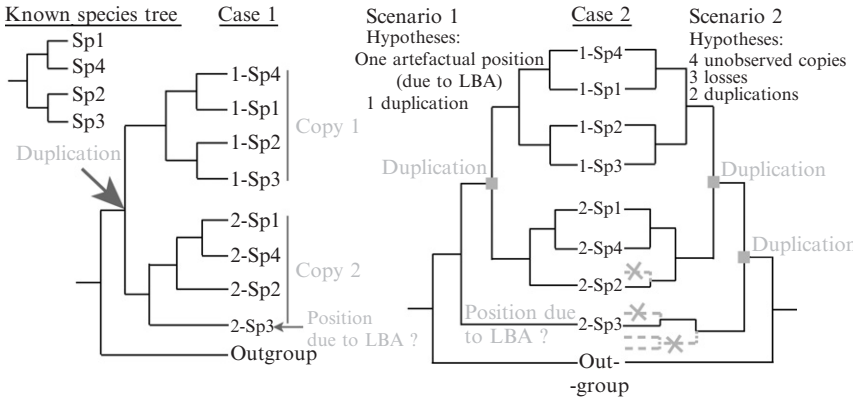

Figure 30.4 Two cases of gene trees and their interpretation in the light of a known species tree. Gray text and lines denote hypotheses. The relative position in the tree of duplications and species divergences allows the inference of a relative timescale for their occurrence. For example, in case 1, gene duplication occurred after the split from the outgroup but before the four species separated from one another. In case 2, scenario 1 should be prefered, as it is more parsimonious. 2-sp3 is not a third paralogue lost multiple times, but a paralogue 2 misplaced by long branch attraction. The position of the copies in the tree allows the inference of orthology and paralogy among them: 1-sp1, 1-sp2, 1-sp3, and 1-sp4 are orthologues, as are 2-sp1, 2-sp2, 2-sp3 and 2-sp4. Any sequence from the copy 1 clade is paralogous to any sequence of the copy 2 clade.

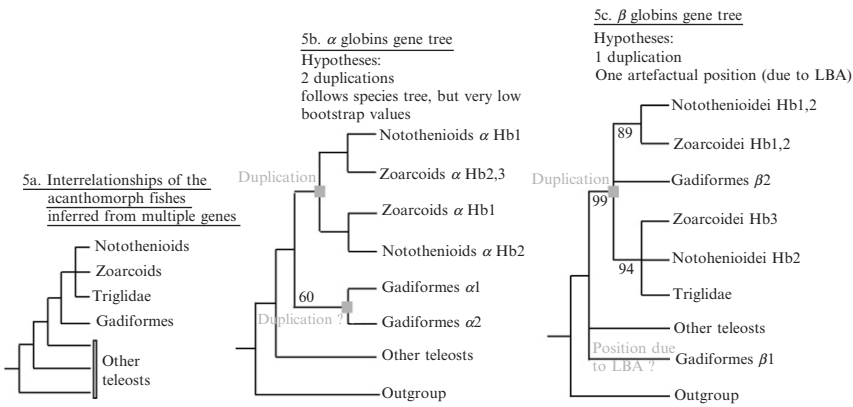

Figure 30.5 Simplified α- and β-globin gene trees for cold-adapted acanthomorph fishes. Adapted from Figures 6 and 7 of Verde *et al.* (2006). Gray text and lines denote hypotheses. The position of *Oncorhynchus mykiss* is unsupported, and the branch has been omitted in this tree for the sake of clarity. The original α-globin tree shows groups of sequences identified as Hb 1, Hb 2, and Hb 3 that form composite clades. A full reevaluation of the nomenclature of the ensemble of fish globins is probably needed. The bootstrap values of the branches of interest have been indicated, except when they are <50.

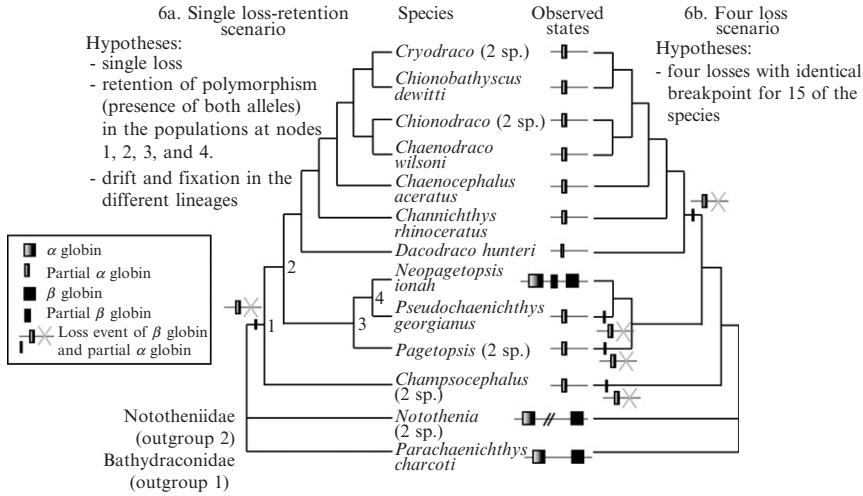

Figure 30.6 Species tree of the notothenioid family Channichthyidae with two different scenarios to explain the distribution of α and β globins in the various species. Adapted from Figures 1 and 2 of Near *et al.* (2006).

Finally, using an adapted inference method is necessary, as reconstruction artefacts cannot be corrected by adding longer sequences or by comparison of several genes. The best inference method remains a highly debated subject, but in such cases it is better not to merely use distance methods like NJ for the analysis and to apply several different approaches.

6.5. Reciprocal illumination

In a multiparalogue, multispecies tree, interpretation of adaptive molecular features is difficult without external information.

Identification of orthologues and paralogues on a gene family tree can greatly benefit from the comparison with a good species-tree approximation. The discrepancies between the two trees will allow for location of which of the nodes represent possible gene duplications, which are more likely due to speciation events, and which are artefacts (see Fig. 30.4). The relative placing of these events in the tree can yield information as to the relative age of the duplication events. Well-dated clade splits can then be used to obtain an estimate of the absolute age of the duplications. A database is available at http://www.timetree.net/ (Hedges *et al.*, 2006), but to date it does not contain information about Actinopterygii. The topology of the tree can be used to detect gene losses (for a more detailed explanation of the identification of paralogues and orthologues, see Cotton [2005]).

As differences between the species and gene trees are crucial in the interpretation, any topological mistake in either one will cause errors in the conclusions drawn from the comparison.

Actually, the use of external information (e.g., knowledge about species tree, duplications) can rarely be omitted in the search for interpretation of species relationships, duplications, and artefacts. All those cannot be deduced in the same inferential process, except in extremely rare cases with a large number of gene copies, where the repetition of the same species interrelationships within each cluster of paralogues gives some reliability to the repeated pattern of species interrelationships.

Species interrelationships are used as solid knowledge to interpret duplication events and artefacts. In case 1 of Figure 30.4, the known species tree allows for interpretation of the position of the copy 2 of species No. 3 as probably artefactual (LBA). In the absence of information about the species interrelationships, it is not possible to decide whether the correct species interrelationships are those exhibited in the part of the gene tree given by copy 1 or by that of copy 2, or even that neither subtree correctly reflects them. Knowledge of species interrelationships often leads to the determination of the number of putative duplication events. The known species tree of figure 30.4 allows to choose the most parsimonious alternative between the two possible scenarii. Scenario 1 implies only one duplication and one LBA, copy losses. Without this knowledge, a third interpretation, as the scenario 1. It implies one duplication (the same as in scenario 1) and one horizontal gene transfer (copy 1 of species No. 2 being transferred into species No. 3).

This strategy has been used by Verde *et al.* (2006) to interpret the discrepancy between the α-globin and the β-globin trees of cold-adapted fish. The position of gadids with regard to zoarcoids and notothenioids is known (see Fig. 30.5A), and allows for hypothesizing that the basal position of the $\beta 1$ sequences of Arctic gadids in the β-globin tree (Fig. 30.5C) is probably artefactual, whereas the α-globin tree recovers mostly the species tree plus a few duplications (Fig. 30.5B). It is more parsimonious to consider the position of the Arctic gadid $\beta 1$-globin sequences as an LBA artefact than a new paralogue-gene cluster not observable in all other fishes (similar to case 2, scenario 1 of Fig. 30.4). LBA was then interpreted as an effect of the extreme perturbation of the available mutational space in gadid $\beta 1$-globin sequences, possibly due to the variability of thermal conditions met by these migratory Arctic fish in comparison with the thermal stability in the lifestyle of zoarcoids and notothenioids, two groups that display unperturbed phylogenetic signals in β sequences.

Conversely, previous knowledge about duplication events can be successfully inserted into the interpretation process to help the inference of species interrelationships. Classical studies on the amino acid sequences of α and β globins exemplify this procedure for other groups (Goodman *et al.*, 1987).

6.6. Reconstruction of character states at the nodes

The inference of character states at each node of the species tree is an essential step in reconstituting the evolution of a character of interest. This character can be a single position of the alignment crucial to the function of the protein or a complex character, such as the presence or absence of a functional motif of several amino acid residues, or even of a whole gene or gene region. Many computer programs allow for such a reconstruction (e.g. Mesquite [Maddison and Maddison, 2006], MacClade [http://macclade.org/index.html], Winclada [Nixon, 1999b], Mr. Bayes 3.1 [Ronquist and Huelsenbeck, 2003]). An extension of reconstruction of ancestral state is the inference of whole ancestral sequences. Such sequences can then be produced in the laboratory and fully tested as it would be possible for present-day sequences (for a review of the methodology and applications, see Chang *et al.*, 2005). Complex characters (e.g., presence or absence of a whole gene copy or active site) must be coded so that the programs can analyze them.

6.7. Alternative reconstructions and interpretations

While mapping characters on a tree might be a simple task, there are often several alternative scenarios to explain the character-state distribution on the tree; examples are shown in Figures 30.4 and 30.5. In some cases, one of the scenarios is much better supported or more parsimonious than its alternatives (Figs. 30.4 and 30.5), but discussion of the others is still of interest. In other cases, several different character mappings have the same support. This is especially frequent when the species tree is not fully resolved, and the various alternatives must all be considered and discussed. Additional background knowledge about molecular evolution can be valuable for selecting scenarios (e.g., the acquisition of exactly the same sequence twice independently might be considered as less likely than two independent losses). The example of the "fossil" α and β globins in *Neopagetopsis ionah* (Near *et al.*, 2006) highlights most of the approaches and tests that can be used in such a case. Two different scenarios have been retained (see Fig. 30.6). One relies on four losses with identical breakpoints of the partial α and the β globins, an unlikely occurrence. The other one hypothesizes a single loss, with retention of ancestral polymorphism in the population through several speciation events. This might seem far-fetched, but the divergences are recent, and several other such instances of retention of molecular polymorphism are known in the channichthyids (Clément *et al.*, 1998), making it actually more interesting than the alternative hypothesis (for the tests and the discussion, see Near *et al.*, 2006). This study also shows an example of the additional information gene trees can bring when used in conjunction with species trees. In this case, a gene tree of the remnants of the α and

a second one of the β globins allowed to trace the origin of the two β-globin copies. These two copies have a very different origin and can probably be explained by an ancient introgression event involving a notothenioid and a channichthyid (see Near *et al.*, 2006).

6.8. Adaptations and disadaptations

It is often difficult to interpret whether a change in sequences is due to adaptation or to phylogenetic innovation, or even both, as they are not mutually exclusive (the adaptation can be a derived character shared by the members of a clade). Adaptive sequence features are sequence changes that can be correlated with some external environmental parameters or lifestyle on to a phylogenetic tree: they correspond to derived character states coupled with external information. The suspicion of being adaptive is even stronger when the same functional and/or environmental information is correlated several times on to a tree with multiple gain of the same derived state. For instance, antifreeze proteins have been gained independently twice in the teleostean tree, in two nonrelated groups (gadids and notothenioids), both of which experience subzero temperatures during part of their lives.

However, for a conclusion about the adaptive nature of the fixation of a character state to be warranted, it must be checked that the putative adaptation is not also fixed in closely related groups not living under the conditions where the adaptation is beneficial. Such groups might not be available, but additional data should be collected to include them whenever possible. For instance, *G. morhua* was included in the study from Verde *et al.* (2006) to provide a noncold-adapted reference within Gadiformes. To make a conclusion on detection of adaptive features, it is useful to keep in mind that correlation is not causation and might only result from the scarcity of groups where data are available. While it might be possible to show that a feature is more largely shared through a clade, even among species with different lifestyles, the lack of such data does not prove that it arose as an adaptation, because we do not have access to the early history of the group.

It is also noteworthy that the notion of adaptation includes the idea that the derived character state is more efficient in the environment than the original (primitive) state of the character. Baum and Larson (1991) proposed the complementary notion of disadaptation, which corresponds to loss of efficiency. Disadaptation can be shown experimentally; the derived character state is then less efficient in the environment than the original one. For instance, icefishes have lost their cardiac myoglobin (Mb) four times, as shown in *Champsocephalus gunnari*, *Pagetopsis*, *Dacodraco hunteri*, and *Chaenocephalus aceratus* (Sidell *et al.*, 1997). These multiple losses favored the hypothesis that cardiac Mb is useless in the whole icefish family Channichthyidae. Acierno *et al.* (1997) have shown that the loss of Mb in the heart of the Antarctic icefish *C. aceratus* corresponds to a disadaptation. They

selectively poisoned cardiac Mb in *Chionodraco myersi*, the sister group living in the same habitats and having natural functional cardiac Mb, and recorded a significant decrease in cardiac output. The Mb is therefore still of use in the heart of these icefishes and had to be compensated for in the groups where it was lost (Montgomery and Clements, 2000; O'Brien and Sidell, 2000). The former conclusion that *C. aceratus* does not have cardiac Mb because it became useless in its ancestors living in cold and oxygen-rich waters is therefore an oversimplification and possibly incorrect.

7. Bringing Phylogenetics and Structural Analysis Together

Understanding the relationships between sequence information and protein function is one of the most challenging goals in structural biology. The inference of the protein functional features requires a multidisciplinary approach, including molecular phylogenetics, structural analysis, and protein engineering. While multiple sequence alignments can be used for reconstructing the phylogeny of protein families, the fold-recognition approach allows for establishing of the relationships with experimentally solved homologous protein structures to be used as a guide in homology modeling. With the aid of suitable bioinformatic tools, it is possible to identify the amino acid sites that are most likely responsible of the functional divergence of different protein families. Mapping these residues on the structural model can be used to establish how relevant they are in determining diversified biochemical and catalytic properties. Finally, site-directed mutagenesis can provide mutant proteins for experimentally testing functional divergence.

7.1. Identification of structural motifs

Structural alignments use information available on the secondary and tertiary structure of proteins. These methods can be used to compare two or more sequences, provided the corresponding structures are known. Because protein structures are more evolutionarily conserved than amino acid sequences (Chothia and Lesk, 1986), structural alignments are usually reliable, especially when comparing distantly related sequences. The simplest way to approach the problem of globin-structure prediction relies on the identification of the sequence regions forming α-helices and strands of β-sheets without inferring the three-dimensional structures of these regions. Several secondary-structure prediction methods are available at the site of the ExPASy tools (http://www.expasy.ch/tools/), including, among others, the powerful hierarchical neural network (HNN) method (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html).

7.2. Molecular modeling

A widely used method for the prediction of a three-dimensional structure of globins starting from known structures of globins is based on homology modeling (Blundell *et al.*, 1987; Fetrow and Bryant, 1993; Johnson *et al.*, 1994). Homology-modeling methods require known protein structures sharing homology to the query sequence to be used as templates and alignment of the query sequence onto the template sequence. Template selection can be achieved by searching in suitable databases by FASTA and BLAST algorithms, choosing hits with low E-values. The quality of the structural model strongly depends on how much the target is related to the template. The entire process of homology modeling includes the selection of a suitable template as a first step, followed by target-template alignment, model-construction, and model-assessment steps. Alignment is critical, because a bad alignment can invalidate the final model. This procedure is now simplified with the development of programs such as Biology Workbench, available at the San Diego Supercomputer Center (http://workbench.sdsc.edu/). A database containing a large number of structures is ModBase (http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi), created by Sali and Blundell (1993), using a program that builds models based on satisfaction of spatial restraints identified from the alignments of homologues of known structure. These restraints are then applied to the unknown sequence. Restraints include distances between α carbon atoms, and other distances within the main-chain, and main-chain and side-chain dihedral angles. Routines to satisfy the restraints optimally include conjugate gradient minimization and molecular dynamics with simulated annealing.

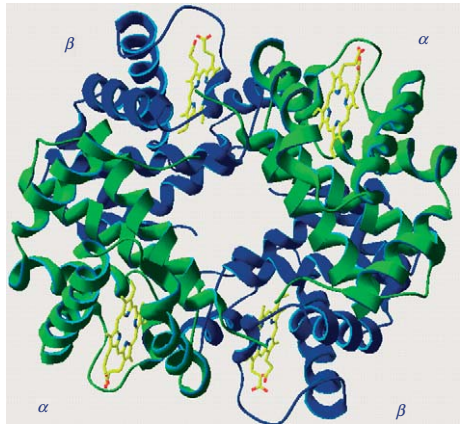

Figure 30.7 Structure of *P. urvillii* Hb 1 (Barbiero, unpublished), built by homology modeling using the program MODELER.

There are several methods for homology modeling. A theoretical model of Hb 1 from the non-Antarctic notothenioid *Pseudaphritis urvillii* is shown in Figure 30.7 as an example. The model (Barbiero, unpublished) is constructed using the crystal structures (Mazzarella *et al.*, 2006a) of *Trematomus bernacchii* Hb (Protein Data Bank, codes 1H8F at pH 6.2 and 1H8D at pH 8.4, respectively) and *T. newnesi* Hb C (Mazzarella *et al.*, 2006b) (Protein Data Bank, code 2AA1). Among notothenioids, *T. bernacchii* Hb has the highest sequence identity with *P. urvillii* Hb 1, at 79 and 82% for the α and β chains, respectively (Verde *et al.*, 2004). The model was built by the MODELER program (Sali and Blundell, 1993) implemented in InsightII (Accelrys Inc). The stereochemistry of the final model was assessed using PROCHECK (Laskowski *et al.*, 1993) and Whatcheck (Hooft *et al.*, 1996).

It is also possible to generate reliable models by homology-modeling tools available on the Web. An automated comparative protein-modeling server is SwissModel (http://swissmodel.expasy.org//SWISS-MODEL.html); it is linked to the software Deep View-Swiss Pdb-Viewer, which provides a friendly interface for analyzing several proteins at the same time. 3DCrunch (http://swissmodel.expasy.org//SM_3DCrunch.html) is aimed at submitting entries from the sequence database directly to SwissModel.

Another fold-recognition method that can be usefully applied in globin-structure prediction is threading. Using a globin as a query sequence, it is possible to predict the protein structure starting from sequence information by comparison with other globins of known structure. This method imposes the query sequence to assume every known protein fold and estimates a scoring function that determines the suitability of the sequence for that particular fold. One of the most promising versions of the method is implemented in the software ROSETTA (Simons *et al.*, 1999) (http://www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php). ROSETTA predicts protein structure by combining the structures of individual fragments inferred through comparison with known structures. The conformational space defined by these fragments is then searched using a Monte Carlo procedure with an energy function that favors compact structures with paired β-sheets and buried hydrophobic residues. The structures provided by such a search are grouped, and the centers of the largest groups are granted as reliable predictions of the target structure.

7.3. Functional divergence

The great similarity of the amino acid sequences and three-dimensional folding shared by vertebrate globins is indicative of their common origin from the same ancestral gene. The evolutionary history of the globin family is characterized by a series of gene-duplication events. The first duplication was probably responsible of the divergence between two functionally distinct oxygen-binding proteins (i.e., Mb and single-chain Hb).

According to the classical model proposed by Ohno (1970), after the occurrence of a gene-duplication event, one of the two gene copies can freely accumulate deleterious mutations that transform it into a pseudogene, provided the other duplicate retains the original function. Alternatively, both duplicates can be preserved if one copy acquires a novel function with the other retaining the ancestral function. The existence of two duplicates, one endowed with the oxygen-carrier function and the other with oxygen-deposit function, associated respectively to Hb and Mb, are in perfect agreement with Ohno's theory. Further duplications brought about the transition from single-chain Hb to α- and β-globin families and subsequent globin multiplicity within each family.

At the molecular level, functional constraints of a protein sequence can provide a measure of the importance of its function. This is equivalent to assuming that functionally relevant residues are conserved, whereas functionally divergent residues violate this constraint (Gu, 1999). Functional and structural divergence can be inferred by statistical methods (Gaucher *et al.*, 2002). The method developed by Gu (1999, 2001, 2003) compares protein sequences encoded by two monophyletic gene clusters generated by gene duplication or speciation. It is assumed that an amino acid residue may have two states. In the S_o state, the site has the same evolutionary rate in both clusters, whereas in the state S_1, the rates in the two clusters are different. The probability of a site to be S_1 is given by the coefficient of Type I functional divergence $\{I\}\theta = P\{/I\}(S_1)$. If the null hypothesis $\{I\}\theta\{/I\} = 0$ is rejected, one can assume that functional constraints at some sites have shifted significantly in the two clusters.

Globin multiplicity observed in many fish species was investigated through the pattern of amino acid substitution during evolution by means of the software DIVERGE (Gu and Vander Velden, 2002), available free of charge at http://xgu1.zool.iastate.edu/, which carries out a likelihood ratio test for the rejection of the null hypothesis $\{I\}\theta\{/I\} = 0$, and allows to identify the residues more likely responsible for the functional divergence between two clusters. The posterior probability $P(S_1 \mid X)$ of a site being related to Type I functional divergence, given the observed amino acid pattern X, is used as a criterion for identifying the most relevant sites.

In the example provided here, the input file consists of a multiple alignment of β-globin sequences in CLUSTAL format. Figure 30.8 shows the gene tree inferred by the NJ method with Poisson distance. In this case, the emphasis is put on the two globin clusters generated by gene duplication during the evolution of Antarctic notothenioids.

In order to perform the analysis, the two clusters of major Hbs and minor Hbs were selected on the gene tree (Parisi *et al.*, unpublished). The output file provides the critical residues predicted by DIVERGE. Among the substitutions having posterior probability >0.80, three (β65 E9, β68

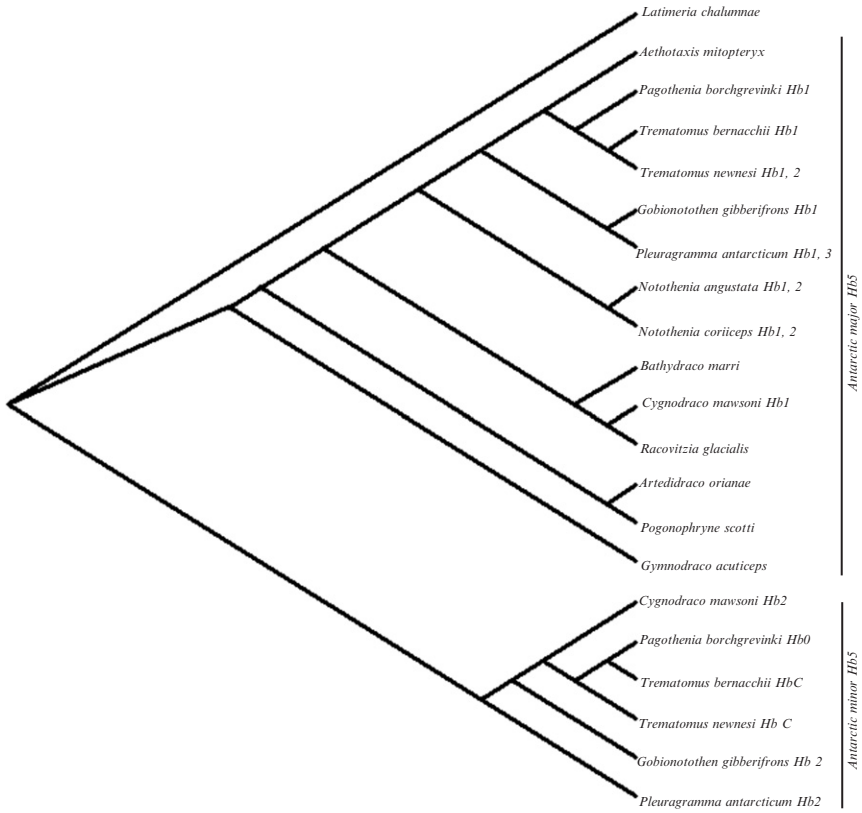

Figure 30.8 Phylogenetic tree of major and minor Antarctic β globins. The tree has been inferred by the NJ method implemented in software DIVERGE and rooted on the sequence of the β globin of *L. chalumnae*. *Notothenia angustata* is a temperate notothenioid.

E12 and β75 E19) occur in helix E, one (β90 F6) in helix F, and one (β121 GH4) in the interhelix region GH (Fig. 30.9).

Another type of functional divergence is Type II, characterized by the site-specific shift of residue properties (Gu, 2006), in contrast to the site-specific shift of evolutionary rate, typical of Type I. Type II functional divergence involves a radical shift of the amino acid properties, such as transition from hydrophilic to hydrophobic character at a homologous site. An example is given by the presence of Pro at a site in one of two hypothetical clusters generated by the gene-duplication event and by the presence of Tyr residue at the corresponding site in the other cluster. The method for studying Type II functional divergence has been implemented in the software DIVERGE2, available at http://xgu.gdcb.iastate.edu.

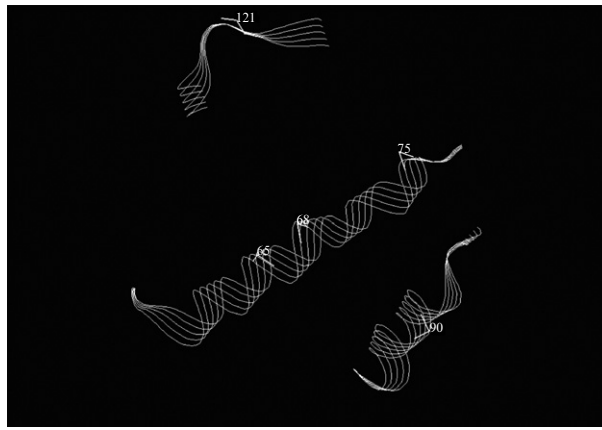

Figure 30.9 Sites involved in Type I functional divergence between major and minor β globins of Antarctic notothenioids. The sites with posterior probability >0.80 are positioned on the structure of helices E and F, and on the interhelix region GH. The three-dimensional model has been built with the software SwissPdbViewer using the atomic coordinates of *T. bernacchii* Hb (Mazzarella *et al.*, 2006a).

In contrast to Type I functional divergence, major and minor Antarctic β globins show no statistically significant Type II functional divergence.

As drastic changes in biochemical properties can often involve the substitution of few amino acid residues, inspection of the sites with the highest value of the posterior ratio score may be useful. Although functional divergence may be very effective to detect residues that contribute to functional-structural diversification, these residues must be only considered as putative candidates for further experiments. Their effectiveness needs to be ascertained with sound experimental data, by coupling the model with additional structural and biochemical information.

8. General Remarks

The joint use of species and gene trees can markedly improve the interpretation of the history of the genes and of adaptations, and allows for avoiding a number of pitfalls in the conclusions. Good achievement of phylogenetic reconstruction in general, as well as integration of the recent progress in the knowledge about species interrelationships, allows for full exploitation of new results about gene and protein structure and function, as they can then be placed in a larger interpretative frame.

Detection of site-specific changes in evolutionary rate can provide a useful tool for predicting the amino acid residues responsible for Type I functional divergence of two monophyletic clusters generated by gene

duplication. The most relevant sites (i.e., those with higher values of posterior probability) can be mapped on the three-dimensional protein structure. In combination with methods aimed at the identification of site-specific shift of the amino acid properties (e.g., Type II functional divergence), this approach offers the opportunity to link sequence and functional/structural divergence.

ACKNOWLEDGMENTS

This study is financially supported by the Italian National Programme for Antarctic Research (PNRA). It is in the framework of the Evolution and Biodiversity in the Antarctic (EBA) program, endorsed by the Scientific Committee on Antarctic Research (SCAR).

REFERENCES

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21,** 2104–2105.

Acierno, R., Agnisola, C., Tota, B., and Sidell, B. D. (1997). Myoglobin enhances cardiac performance in Antarctic icefish species that express the pigment. *Am. J. Physiol.* **273,** R100–R106.

Antonini, E., Rossi-Bernardi, L., and Chiancone, E. (1981). Hemoglobins. *Methods Enzymol.* **76,** 874.

Bargelloni, L., Marcato, S., and Patarnello, T. (1998). Antarctic fish hemoglobins: Evidence for adaptive evolution at subzero temperature. *Proc. Natl. Acad. Sci. USA* **95,** 8670–8675.

Barriel, V. (1994). Phylogenies moleculaires et insertions-deletions de nucleotides. *C. R. Acad. Sci. B* **317,** 693–701.

Baum, D. A., and Larson, A. (1991). Adaptation reviewed: A phylogenetic methodology for studying character macroevolution. *Syst. Zool.* **40,** 1–18.

Berenbrink, M., Koldkjaer, P., Kepp, O., and Cossins, A. R. (2005). Evolution of oxygen secretion in fishes and the emergence of a complex physiological system. *Science* **307,** 1752–1757.

Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., and Thornton, J. M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326,** 347–352.

Brauer, A. W., Oman, C. L., and Margolies, M. N. (1984). Use of *o*-phthalaldehyde to reduce background during automated Edman degradation. *Anal. Biochem.* **137,** 134–142.

Bremer, K. (1994). Branch support and tree stability. *Cladistics* **10,** 295–304.

Brittain, T. (2005). Root effect hemoglobins. *J. Inorg. Biochem.* **99,** 120–129.

Chang, B. S. W., and Campbell, D. L. (2000). Bias in phylogenetic reconstruction of vertebrate rhodopsin sequences. *Mol. Biol. Evol.* **17,** 1220–1231.

Chang, B. S. W., Ugalde, J. A., and Matz, M. V. (2005). Applications of ancestral protein reconstruction in understanding protein function: GFP-like proteins. *Methods Enzymol.* **395,** 652–670.

Chen, W.-J., Bonillo, C., and Lecointre, G. (2003). Repeatability as a criterion of reliability of new clades in the acanthomorph (Teleostei) radiation. *Syst. Biol.* **26,** 262–288.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159.

Chothia, C., and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5,** 823–826.

Clément, O., Ozouf-Costaz, C., Lecointre, G., and Berrebi, P. (1998). Allozymic polymorphism and the phylogeny of family Channichthyidae. *In* "Fishes of Antarctica, a biological overview" (G. di Prisco, E. Pisano, and A. Clarke, eds.) pp. 299–309. Springer-Verlag, Berlin.

Cotton, J. A. (2005). Analytical methods for detecting paralogy in molecular datasets. *Methods Enzymol.* **395,** 700–724.

De Queiroz, A. (1993). For consensus (sometimes). *Syst. Biol.* **42,** 368–372.

DeSalle, R. (2005). Animal phylogenomics: Multiple interspecific genome comparisons. *Methods Enzymol.* **395,** 104–133.

Dettaï, A., and Lecointre, G. (2004). In search of the Notothenioid relatives. *Antarctic Sci.* **16,** 71–85.

Dettaï, A., and Lecointre, G. (2005). Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. *C. R. Acad. Sci. B* **328,** 674–689.

Dingerkus, G., and Howell, W. M. (1976). Karyotypic analysis and evidence of tetraploidy in the North American paddlefish, *Polyodon spathula*. *Science* **194,** 842–844.

di Prisco, G., Eastman, J. T., Giordano, D., Parisi, E., and Verde, C. (2007). Biogeography and adaptation of Notothenioid fish: Hemoglobin function and globin-gene evolution. *Gene* (in press).

Everse, J., Vandegriff, K. D., and Winslow, R. M. (1994). Hemoglobins part B: Biochemical and analytical methods. *Methods Enzymol.* **231,** 725.

Felsenstein, J. (2004). "Inferring phylogenies." Sunderland, MA: Sinauer, Sunderland, MA.

Fetrow, J. S., and Bryant, S. H. (1993). New programs for protein tertiary structure prediction. *Bio/Technology* **11,** 479–484.

Filleul, A., and Lavoué, S. (2001). Basal teleosts and the question of elopomorph monophyly: Morphological and molecular approaches. *C. R. Acad. Sci. B.* **324,** 393–399.

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Syst. Zool.* **19,** 99–113.

Force, A., Lynch, M., Pickett, F. B., Amores, A., and Yan, Y. L. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545.

Friedman, M., Krull, L. H., and Cavins, J. F. (1970). The chromatographic determination of cystine and cysteine residues in proteins as *S*-β-(4-pyridylethyl) cysteine. *J. Biol. Chem.* **245,** 3868–3871.

Gascuel, O., and Steel, M. (2006). Neighbor-joining revealed. *Mol. Biol. Evol.* **23,** 1997–2000.

Gaucher, E. A., Gu, X., Miyamoto, M., and Benner, S. (2002). Predicting functional divergence in protein evolution by site-specific rate shifts. *Trend Biochem. Sci.* **27,** 315–321.

Giordano, D., Grassi, L., Parisi, E., Bargelloni, L., di Prisco, G., and Verde, C. (2006). Embryonic *b*-globin in the non-Antarctic notothenioid fish *Cottoperca gobio* (Bovichtidae). *Polar Biol.* **30,** 75–82.

Goloboff, P., Farris, S., and Nixon, K. (2000). TNT (tree analysis using new technology) (BETA) Tucumán, Argentina: Published by the authors, Tucumán, Argentina.

Goodman, M., Miyamoto, M. M., and Czelusniak, J. (1987). Pattern and process in vertebrate phylogeny revealed by coevolution of molecules and morphologies. *In* "Molecules and morphology in evolution: Conflict or compromise?" (C. Patterson, ed.), pp. 141–176. Cambridge University Press, New York.

Graybeal, A. (1994). Evaluating the phylogenetic utility of genes: A search for genes informative about deep divergences among vertebrates. *Syst. Biol.* **43,** 174–193.

Graybeal, A. (1998). Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47,** 9–17.

Gu, X. (1999). Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol.* **16,** 1664–1674.

Gu, X. (2001). Maximum likelihood approach for gene family evolution under functional divergence. *Mol. Biol. Evol.* **18,** 453–464.
Gu, X. (2003). Functional divergence in protein (family) sequence evolution. *Genetica* **118,** 133–141.
Gu, X. (2006). A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol. Biol. Evol.* **23,** 1937–1945.
Gu, X., and Vander Velden, K. (2002). DIVERGE: Phylogeny-bases analysis for functional-structural divergence of a protein family. *Bioinformatics* **18,** 500–501.
Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704.
Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41,** 95–98.
Hashiguchi, Y., and Nishida, M. (2006). Evolution and origin of vomeronasal-type odorant receptor gene repertoire in fishes. *BMC Evol. Biol.* **6,** 76.
Hedges, S. B., Dudley, J., and Kumar, S. (2006). TimeTree: A public knowledge-base of divergence times among organisms. *Bioinf.* **22,** 2971–2972.
Hendy, M. D., and Penny, D. (1989). A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38,** 297–309.
Higgins, D. G., and Sharp, P. M. (1988). CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73,** 237–244.
Hillis, D. M. (1987). Molecular versus morphological approaches to systematics. *Ann. Rev. Ecol. Syst.* **18,** 23–42.
Hillis, D. M., Pollock, D. D., McGuire, J. A., and Zwickl, D. J. (2003). Is sparse sampling a problem for phylogenetic inference? *Syst. Biol.* **52,** 124–126.
Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996). Errors in protein structures. *Nature* **381,** 272.
Hurles, M. (2004). Gene duplication: The genomic trade in spare parts. *PLoS Biol.* **2,** e206.
Hurley, I. A., Lockridge Mueller, R., Dunn, K. A., Schmidt, E. J., Friedman, M., Ho, R. K., Prince, V. E., Yang, Z., Thomas, M. G., and Coates, M. I. (2007). A new time-scale for ray finned fish evolution. *Proc. R. Soc. B.* **274,** 489–498.
Inoue, J. G., Miya, M., Tsukamoto, K., and Nishida, M. (2003). Basal actinopterygian relationships: A mitogenomic perspective on the phylogeny of the ''ancient fish.'' *Mol. Phylogenet. Evol.* **26,** 110–120.
Ishiguro, N. B., Miya, M., and Nishida, M. (2003). Basal euteleostean relationships: A mitogenomic perspective on the phylogenetic reality of the ''Protacanthopterygii.'' *Mol. Phylogenet. Evol.* **27,** 476–488.
Janvier, P. (1996). ''Early vertebrates.'' : Clarendon University Press, Oxford, UK.
Johnson, M. S., Srinivasan, N., Sowdhamini, R., and Blundell, T. L. (1994). Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.* **29,** 1–68.
Kikugawa, K., Katoh, K., Kuraku, S., Sakurai, H., Ishida, O., Iwabe, N., and Miyata, T. (2004). Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Biol.* **2,** 3.
Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). *Syst. Zool.* **38,** 7–25.
Kuo, M. W., Postlethwait, J., Lee, W. C., Lou, S. W., Chan, W. K., and Chung, B. C. (2005). Gene duplication, gene loss and evolution of expression domains in the vertebrate nuclear receptor NR5A (Ftz-F1) family. *Biochem. J.* **389,** 19–26.
Landon, M. (1977). Cleavage at aspartyl-prolyl bonds. *Methods Enzymol.* **47,** 145–149.
Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7,** S4.
Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thorntorn, J. M. (1993). PROCHECK: A program to check the stereochemical quality of proteins structures. *J. Appl. Crystallogr.* **26,** 283–291.

Lê, H. L. V., Lecointre, G., and Perasso, R. (1993). A 28S rRNA based phylogeny of the gnathostomes: First steps in the analysis of conflict and congruence with morphologically based cladograms. *Mol. Phylogenet. Evol.* **2,** 31–51.
Lecointre, G., Philippe, H., Van Le, H. L., and Le Guyader, H. (1993). Species sampling has a major impact on phylogenetic inference. *Mol. Phylogenet. Evol.* **2,** 205–224.
Long, M., Betran, E., Thornton, K., and Wang, W. (2003). The origin of new genes: Glimpses from the new and the old. *Nature Reviews Genet.* **4,** 865–875.
Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290,** 1151–1155.
Maddison, M. P. (1997). Gene trees in species trees. *Syst. Biol.* **46,** 523–536.
Maddison, W. P., and Maddison, D. R. (2006). Mesquite: A modular system for evolutionary analysis (http://mesquiteproject.org).
Maruyama, K., Yasumasu, S., and Iuchi, I. (2004). Evolution of globin genes of the medaka *Oryzias latipes* (Euteleostei; Beloniformes; Oryziinae). *Mech. Dev.* **121,** 753–769.
Mathews, S. (2005). Analytical methods for studying the evolution of paralogs using duplicate gene datasets. *Methods Enzymol.* **395,** 724–745.
Mazzarella, L., Vergara, A., Vitagliano, L., Merlino, A., Bonomi, G., Scala, S., Verde, C., and di, Prisco. (2006a). High resolution crystal structure of deoxy haemoglobin from *Trematomus bernacchii* at different pH values: The role of histidine residues in modulating the strength of the Root effect. *Proteins Str. Funct. Bioinf.* **65,** 490–498.
Mazzarella, L., Bonomi, G., Lubrano, M. C., Merlino, A., Riccio, A., Vergara, A., Vitagliano, L., Verde, C., and di Prisco, G. (2006b). Minimal structural requirements for Root effect: Crystal structure of the cathodic hemoglobin isolated from the Antarctic fish *Trematomus newnesi*. *Proteins Str. Funct. Bioinf.* **62,** 316–321.
Miya, M., Satoh, T. P., and Nishida, M. (2005). The phylogenetic position of toadfishes (order Batrachoidiformes) in the higher ray-finned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. *Biol. J. Linn. Soc.* **85,** 289–306.
Miya, M., Takeshima, H., Endo, H., Ishiguro, N. B., Inoue, J. G., Mukai, T., Satoh, T. P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S. M., and Nishida, M. (2003). Major patterns of higher teleostean phylogenies: A new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **26,** 121–138.
Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44,** 64–76.
Montgomery, J., and Clements, K. (2000). Disaptation and recovery in the evolution of Antarctic fishes. *Trends Ecol. Evol.* **15,** 267–271.
Near, T. J., Parker, S. K., and Detrich, H. W., III (2006). A genomic fossil reveals key steps in hemoglobin loss by the Antarctic fishes. *Mol. Biol. Evol.* **23,** 2008–2016.
Nei, M., and Kumar, S. (2000). "Molecular phylogenetics and evolution." New York: Oxford University Press, New York.
Nelson, J. S. (2006). "Fishes of the world." Hoboken, NJ: John Wiley and Sons, Hoboken, NJ.
Nixon, K. C. (1999a). The parsimony ratchet: A new method for rapid parsimony analysis. *Cladistics* **15,** 407–414.
Nixon, K. C. (1999b). Winclada (BETA) Ver. 0.9.9 Published by the author. Ithaca, NY.
Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302,** 205–217.
O'Brien, K. M., and Sidell, B. D. (2000). The interplay among cardiac ultrastructure, metabolism and the expression of oxygen-binding proteins in Antarctic fishes. *J. Exp. Biol.* **203,** 1287–1297.
Ohno, S. (1970). "Evolution by gene duplication." Berlin: Springer-Verlag, Berlin.
O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G., and Notredame, C. (2004). 3DCoffee: Combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.* **340,** 385–395.

Page, R. D. M., and Holmes, E. C. (1998). "Molecular evolution: A phylogenetic approach," Abingdon, UK: Blackwell Science, Abingdon, UK.
Philippe, H., and Douzery, E. (1994). The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. *J. Mam. Evol.* **2,** 133–152.
Posada, D., and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53,** 793–808.
Posada, D., and Crandall, K. A. (1998). ModelTest: Testing the model of DNA substitution. *Bioinf.* **14,** 817–818.
Prince, V. E., and Pickett, F. B. (2002). Splitting pairs: The diverging fates of duplicated genes. *Nature Rev. Genet.* **3,** 827–837.
Riggs, A. (1988). The Bohr effect. *Annu. Rev. Physiol.* **50,** 181–204.
Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P. L., Zelus, D., Hughes, S., and Laudet, V. (2001). Euteleost fish genomes are characterized by expansion of gene families. *Genome Res.* **11,** 781–788.
Ronquist, F., and Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinf.* **19,** 1572–1574.
Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstruction of phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.
Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815.
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: A laboratory manual 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sidell, B. D., and O'Brien, K. M. (2006). When bad thing happen to good fish: The loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J. Exp. Biol.* **209,** 1791–1802.
Sidell, B. D., Vayda, M. E., Small, D. J., Moylan, T. J., Londraville, R. L., Yuan, M.-L., Rodnick, K. J., Eppley, Z. A., and Costello, L. (1997). Variation in the expression of myoglobin among species of the Antarctic icefishes. *Proc. Natl. Acad. Sci. USA* **94,** 3420–3424.
Sikes, D. S., and Lewis, P. O. (2001). PAUPRat: PAUP* implementation of the parsimony ratchet (Beta), Version 1 Storrs, CT: Distributed by the authors, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT.
Simons, K. T., Bonneau, R., Ruczinski, I., and Baker, D. (1999). Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* **3,** 171–176.
Smith, W. L., and Wheeler, W. C. (2004). Polyphyly of the mail-cheeked fishes (Teleostei: Scorpaeniformes): Evidence from mitochondrial and nuclear sequence data. *Mol. Phylogenet. Evol.* **32,** 627–646.
Smith, W. L., and Wheeler, W. C. (2006). Venom evolution widespread in fishes: A phylogenetic road map for the bioprospecting of piscine venoms. *J. Hered.* **97,** 206–217.
Stiassny, M. L. J., Parenti, L. R., and Johnson, G. D. (1996). "The interrelationships of fishes." San Diego: Academic Press, San Diego.
Swofford, D. L. (2001). "PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)." Sunderland, MA: Version 4. Sinauer Associates, Sunderland, MA.
Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z., and Klein, J. (2003). Molecular phylogeny of early vertebrates: Monophyly of the agnathans as revealed by sequences of 35 genes. *Mol. Biol. Evol.* **20,** 287–292.
Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z., Takahata, N., and Klein, J. (2004). The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of 44 nuclear genes. *Mol. Biol. Evol.* **21,** 1512–1524.
Tamburrini, M., Brancaccio, A., Ippoliti, R., and di Prisco, G. (1992). The amino acid sequence and oxygen-binding properties of the single hemoglobin of the cold-adapted Antarctic teleost *Gymnodraco acuticeps*. *Arch. Biochem. Biophys.* **292,** 295–302.

Tamburrini, M., D'Avino, R., Fago, A., Carratore, V., Kunzmann, A., and di Prisco, G. (1996). The unique hemoglobin system of *Pleuragramma antarcticum*, an Antarctic migratory teleost. Structure and function of the three components. *J. Biol. Chem.* **271,** 23780–23785.

Taylor, J. S., Van de Peer, Y., Braasch, I., and Meyer, A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356,** 1661–1679.

Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22.000 species of ray-finned fish. *Genome Res.* **13,** 382–390.

Thompson, J. D., Gibson,T. J., Plewniak,F., Jeanmougin, F., and Higgins, D. G. (1997). The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24,** 4876–4882.

Uyeno, T., and Smith, G. R. (1972). Tetraploid origin of the karyotype of catostomid fishes. *Science* **175,** 644–646.

Van de Peer, Y., Taylor, J. S., Joseph, J., and Meyer, A. (2002). Wanda: A database of duplicated fish genes. *Nucleic Acids Res.* **30,** 109–112.

Venkatesh, B., Erdmann, M. V., and Brenner, S. (2001). Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc. Nat. Acad. Sci. USA* **98,** 11382–11387.

Verde, C., Balestrieri, M., de Pascale, D., Pagnozzi, D., Lecointre, G., and di Prisco, G. (2006). The oxygen transport system in three species of the boreal fish family Gadidae. *J. Biol. Chem.* **283,** 22073–22084.

Verde, C., Howes, B. D., De Rosa, M. C., Raiola, L., Smulevich, G., Williams, R., Giardina, B., Parisi, E., and di Prisco, G. (2004). Structure and function of the Gondwanian hemoglobin of *Pseudaphritis urvillii*, a primitive notothenioid fish of temperate latitudes. *Prot. Sci.* **13,** 2766–2781.

Wiens, J. J. (1998). Combining data sets with different phylogenetic histories. *Syst. Biol.* **47,** 568–581.

Wittenberg, J. B., and Wittenberg, B. A. (1974). The choroid *rete* mirabilis. 1. Oxygen secretion and structure: Comparison with the swimbladder *rete* mirabile. *Biol. Bull.* **146,** 116–136.

Zaragueta-Bagils, R., Lavoué, S., Tillier, A., Bonillo, C., and Lecointre, G. (2002). Assessment of otocephalan and protacanthopterygian concepts in the light of multiple molecular phylogenies. *C. R. Acad. Sci. B* **325,** 1191–1207.