Antarctic Science 16 (1): 000–000 (2004) © Antarctic Science Ltd Printed in the UK DOI: 10.1017/S0954102004

In search of notothenioid (Teleostei) relatives

AGNÈS DETTAÏ and GUILLAUME LECOINTRE*

UMR 7138 CNRS "Systématique, Adaptation, Evolution", Département Systématique et Evolution, Muséum National d'Histoire Naturelle, 43 rue Cuvier 75231 PARIS cedex 05, France
**corresponding author: lecointr@mnhn.fr*

Abstract: Ninety-five percent of the fish species known from the Antarctic continental shelf and upper slope are acanthomorphs, i.e. spiny teleosteans. Notothenioids (suborder Notothenioidei) are acanthomorphs and so is their sister group. Unfortunately, until recently acanthomorph intra-relationships were so poorly known that it was necessary to sample all of this diversity just to search for a single sister group relationship. Using recent advances in acanthomorph molecular phylogenetics, particular properties of separate analyses and a new protocol of dataset combination, we identified a clade that contains the sister group of notothenioids, the Percidae (perches), and a number of relatives. Among these relatives are the Serranidae (sea basses), the genera *Trachinus* (weeverfish), *Chelidonichthys* (gurnard), *Scorpaena* (scorpionfish), and a group composed of the Zoarcoidei (eelpouts) and the Cottoidei (sculpins) with the Gasterosteidae (sticklebacks) as their sister group. Interestingly, that clade contains 88% of the fish species found on the Antarctic continental shelf and upper slope. The interrelationships of its components and their distribution show that the Antarctic benthic fish fauna has at least three origins.

Received 6 May 2003, accepted 8 July 2003

Key words: Acanthomorpha, benthic Antarctic fishes, molecular phylogeny, Notothenioidei, taxonomic congruence

Introduction

The Notothenioidei contains 122 species of Antarctic and sub-Antarctic teleostean fishes, of which 96 are strictly Antarctic (Eastman & Clarke 1998). The suborder represents 46% of the "fish" species living on the Antarctic continental shelf and upper slope, and 90% of the fish biomass. They are members of the Acanthomorpha, a wide group that comprises all teleosts possessing spines in their dorsal and anal fins. As for the rest of the 208 species known on the Antarctic continental shelf and upper continental slope, 31% belong to the family Liparidae, 11% to the Zoarcidae, with a few more "fish" species from other families. According to Eastman & Clarke (1998, table 2), 95% of the species known there are acanthomorphs. Obviously, if one wants to understand the history of colonization of the Southern Ocean by teleosts, acanthomorph phylogeny is indeed relevant to the question.

There is no doubt that notothenioids are acanthomorphs, and therefore that their sister group also belongs to that group. In the December 2002 FishBase (Froese & Pauly 2003) census, the number of species in the acanthomorph clade was 15 344, distributed in 314 families, and representing nearly 60% of extant "fish" diversity. The group is so large and diverse that, for obvious practical reasons, one would like to look for the sister group of notothenioids within a smaller group. Here lies the major problem. The question of the notothenioid origin is difficult to approach, because until very recently there was no reliable group smaller than the whole Acanthomorpha to sample, although Percomorpha, Perciformes and Percoidei had been suggested but the Perciformes have never been clearly defined and are generally considered to be polyphyletic. The Percomorpha of Johnson & Patterson (1993) are defined by a single synapomorphy, and recent molecular phylogenies have repeatedly refuted the validity of this grouping (Chen *et al.* 2000, 2001, 2003, Miya *et al.* 2001, 2003a). The same situation stands for a number of larger clades within acanthomorphs as defined by Johnson & Patterson (1993). As a consequence, finding the sister group of notothenioids requires a huge number of terminal taxa to cover acanthomorph diversity adequately, making it an ambitious, time and money consuming undertaking. From a morphological and anatomical point of view, notothenioid families have been placed among percoids for a long time (Regan 1913, Norman 1937, 1938, 1966, Berg 1940, Andriashev 1964, Lindberg 1971). However, percoids ("perch-like fishes") have never been really defined in terms of common exclusive character states. A number of authors suggested more precise hypotheses of kinship for the Notothenioidei, placing them together with some trachinoid components (weeverfish-like fishes: Berg 1947, Bertin & Arambourg 1958, Gosline 1968, Hastings 1993, Balushkin 2000), or with the Zoarcoidei (eelpouts: Anderson 1984, 1990). Among the polyphyletic trachinoids, the Trichonotidae (sanddivers: Gosline 1968, Hastings 1993), Pinguipedidae (sandperches: Gosline 1968, Pietsch 1989, Hastings 1993, Balushkin 2000) and Cheimarrichthyidae (torrentfishes: Gosline 1968) have been proposed as sister groups of the notothenioids. Blennioidei have also been invoked in the literature (Eastman 1993,

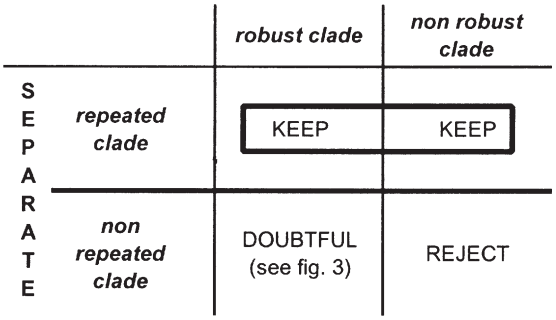

Fig. 1. Summary of the methodological framework used for the assessment of the reliability of the clades. The upper plain line rectangle indicates results considered as reliable. We keep the tree from the simultaneous analysis as the major tree. However, reliability of the clades is taken from their repeatability through separate analyses.

Balushkin 2000), as Gosline included those trachinoid components as well as notothenioids under a wider understanding of the Blennioidei. Most modern authors have left the Notothenioidei in the big perciform bush (Greenwood *et al.* 1966, Nelson 1994).

Wide taxonomic samplings have been carried out for molecular phylogenies within the last few years (Chen *et al.* 2000, 2001, 2003, Miya *et al.* 2001, 2003a, 2003b, Dettaï & Lecointre **submitted**, hereafter referred to as Dettaï & Lecointre). The protocol used to assess reliability of clades in Chen *et al.* (2003) permitted the identification of a reliable candidate for the sister group status: from several genes analysed separately, Chen *et al.* (2003) and Dettaï & Lecointre identified the Percidae (perches) as the sister group of the notothenioids. Though the Trichonotidae and the Pinguipedidae *sensu stricto* have not been sampled yet, molecular studies rejected a number of the aforementioned morphology-based hypotheses. The torrentfish of New Zealand, *Cheimarrichthys* (generally considered as a pinguipedid, but sometimes put in its own family of Cheimarrichthyidae) is not the sister group of the sub-Antarctic and Antarctic Notothenioidei, but is instead closer to sand lances (Ammodytidae). Other trachinoids such as the Chiasmodontidae group with the Scombroidei (mackerels) and the Stromateoidei; zoarcoids (eelpouts) are close relatives of cottoids (sculpins) and gasterosteids (sticklebacks). The blennies sampled to date group with the Gobiesocoidei (clingfishes).

Phylogenetic relationships provided by these molecular analyses are interesting to consider in more detail. Classical candidates for the sister group of notothenioids (Zoarcoidei, Trachinoidei) were rejected because they were repeatedly found with other components. However, several of the classical candidates, with their unexpected sister groups, were placed close in the big acanthomorph tree in recent molecular studies. A larger clade emerged, encompassing some of those classical candidates, as well as some other fishes such as serranids (sea basses), gasterosteids, and scorpaenoids. For convenience, we will refer to this group as "clade X". That wider clade, although present, could not be identified as reliable in the studies of Chen *et al.* (2003) or Dettaï & Lecointre, based on nuclear genes and 12S and 16S mitochondrial rDNAs, because of the criterion of reliability they used. In these studies, a clade was considered to be reliable if it was repeatedly found in analyses of several independent genes. But then the clade was also recovered by Miya *et al.* (2003b) from an important number of mitochondrial genes (excluding 12S and 16S rDNAs), giving it a first confirmation. The aim of the present paper is to show that, if we take into account some unexploited properties of separate phylogenetic analyses, it is possible to propose a reliable clade containing the notothenioid origins, far smaller than that of the entire acanthomorph bush, and more comprehensive than the clade containing only the Percidae and the Notothenioidei.

Material and methods

General principles

When several independent genes are sequenced for a given taxonomic sampling, each gene can be analysed separately (separate analyses), leading to several trees that may or may not be congruent. Alternatively, all the genes can be combined into a single dataset, and the phylogenetic analysis conducted then is called simultaneous analysis (Nixon & Carpenter 1996). As already explained in detail in Chen *et al.* (2003) and Dettaï & Lecointre, the methodology we adopted uses both methods, establishing a tree based upon all available sequence data combined in order to study the history of a trait, to maximize the congruence of the characters and to get an estimate of statistical robustness for each clade. Yet the reliability of each clade is obtained from the recurrence of its presence, whatever its bootstrap proportion, across the separate analyses of each independent gene (for detailed arguements see Chen *et al.* 2003). This methodological framework using both separate and simultaneous analyses (Lecointre & Deleporte 2000) is summarized in Fig. 1.

The difficulties in interpreting multiple molecular phylogenies come, in part, from the possible advantages of simultaneous and disadvantages of separate analyses, (see Miyamoto & Fitch 1995). Simultaneous analyses have two main advantages: they lead to trees that maximize the congruence of all the available characters, and often allow a gain in statistical robustness for clades. But in these analyses it is not possible to differentiate signals caused by a common origin of taxa from non-phylogenetic signals. The risk is then to wrongly identify a robust grouping resulting from positively misleading signals, such as heterogeneity across taxa in base-composition, or in rates of evolution, or from changes of mutational space across taxa at some positions, for phylogenetic interrelationships. The positively misleading signal may be caused by a single

gene, but can impose a "false" topology on the tree resulting from the simultaneous analysis if the other datasets are not sufficiently structured. Since the constraints bringing about the misleading signal may be the result of selective pressures acting on one gene, and the likelihood of similar constraints working in the same way on several independent markers is low, separate analyses of independent genes help in detecting reliable clades. If a clade is found repeatedly, despite the fact that different markers are likely to be subject to different selective pressures, it is reasonable to suppose that this clade must reflect the genealogical history of taxa and is not the product of selection biases. Thus, working on the results of the separate analyses without making an overall consensus tree allows for the assessment of the reliability of clades through their repeatability, regardless of the statistical robustness associated with the clade in each tree.

Such separate phylogenetic analyses stand on the justification of data partition. Much criticism has been made of the assertion that partitions are natural, suggesting instead that they are the product of technical and historical constraints rather than the reflection of some biological properties. We work under the assumption that genes are natural partitions, and that the way homoplasy is accumulated throughout each gene is different so long as the genes are independent. There are several reasons for holding this position. We claim that, if something is known about our data, that knowledge must be explicitly rather than implicitly used. Taking that knowledge into account or not is partly linked to the question of "falsification" and the nature of "tests" in systematics (Sober 1988, ch. 4, Mahner & Bunge 1997, pp. 124, Rieppel 2003), which is beyond the scope of the present paper. In most cases a gene can be considered as a selective unit with peculiar constraints because the protein it codes for is a self-delineated and integrated physical entity. When these constraints are homogenous along the gene, and heterogenous across different genes, they can lead to various effects on tree reconstruction, sometimes in the form of positively misleading signals. For example, the fact that a vertebrate cytochrome *b* has an anti-G bias in all the third codon positions has a biological explanation, and gives the gene particular properties with regard to phylogenetic reconstruction, different from the ones found, for example, in a nuclear ribosomal gene. For instance, rapid saturation at transitions of the first codon positions of the cytochrome *b* gene has a connection with rapid hydrophobic amino acid interchanges in the hydrophobic segments of the protein. Such data biases, e.g. high GC content, can cause convergences among unrelated taxa that will then be clustered in the tree. They can provoke long branches that will also cause the affected taxon or taxa to be moved to a more basal position, out of the clade where it is found from other genes. Although these situations are sometimes easily recognizable, as in the case of a taxon that is included in a group in the trees from most datasets and escapes to a totally unrelated group in the tree from one of the genes, they greatly complicate the recovery of larger groups in their entirety.

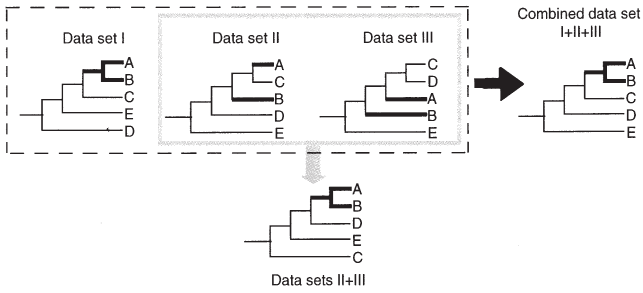

Fig. 2. Research team 1 found the clade A+B in the simultaneous analysis of I+II+III. It was not considered as reliable because it was not repeated in separate analyses of I, II, III. A+B could not be recovered by datasets II and III, possibly because of stochastic errors. A way to test that hypothesis is to conduct the phylogenetic analysis of the partial combination II+III.

However, separate analyses also harbour some pitfalls. Among these is implicit uncontrolled weighting: by comparing trees, we implicitly give an equal weight to sequences of different sizes, applying an uncontrolled weight to each position in the sequence. The protocol we used reduces the impact of this criticism. In any tree we consider of interest only the clades that are repeated in other trees, and those clades do not depend on weights given to each position, as long as the weights are homogenous within a gene (which is the case).

The last point of criticism is the one we would like to explore in the present paper: that separate analyses still bear the risk of stochastic errors. A partition can be too small and contain too little information to enable the recovery of a given clade. If the positions supporting that clade are rare, homoplasy can take over and components of that clade will exhibit unstable positions across trees based on different genes. Collecting only the clades that are repeated is a cautious approach because stochastic errors are not repeated, but this causes the loss of quite a lot of information. But there is a way to extract more information from datasets presenting this kind of stochastic error. As explained before, combination of the datasets should allow signal recovery but combination of all the data deprives us of any way to assess reliability. Yet if partial combinations of the data are also performed, they can get round both problems. Partial (i.e. incomplete) combination produces longer datasets containing subparts with different evolutionary constraints, and are therefore less prone to stochastic errors. And it is still possible to check in a rigorous way for the recurrence of a clade by comparing different combinations that contain no data in common, and are therefore probably independent. An example of this is presented in Fig. 2.

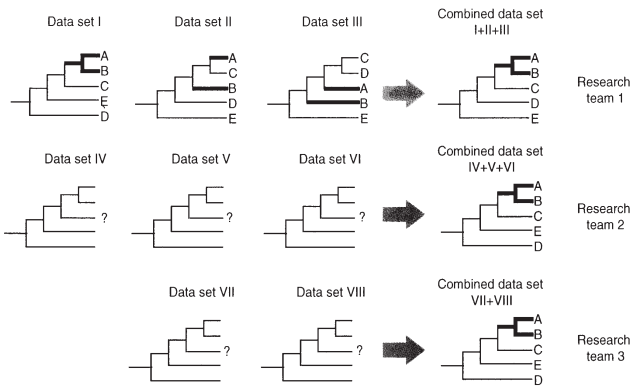

Fig. 3. Theoretical example of the use of taxonomic congruence as a tool to assess reliability, when based on trees generated by different research teams. Each team bases its conclusions on the tree from the combined dataset, so the results of the separate analyses are not always available (hence the question marks in front of the separate analysis trees of the other teams). When considering the results of all teams, the clade A+B is recurrently found and can be considered as reliable. However team 1 could not accept it as reliable on the basis of their results alone.

The robust clade A+B found in the tree from the simultaneous analysis shown in Fig. 2 could be the correct one, but it is found only once across separate analyses because of stochastic errors in the other analyses. Datasets II and III (Fig. 2) were just too small to recover A+B. So the test to decide whether A+B is an artifact due to a positively misleading signal in the gene I (Fig. 2), or a reliable clade not repeated in II and III because of stochastic errors, is the following: If we perform the incomplete combination II+III, providing a dataset that will contain more information, and, as such, might be less prone to stochastic errors. If A+B appears in the resulting tree, the clade can be considered as repeated from independent data. This interpretation is based on the assumption that, if A+B is really due to a bias in I, there is little chance to recover it in the independent II+III. If A+B does not appear from the combination II+III, either A+B can be the result of a bias in I or else II+III can still be too small to recover the correct grouping. The clade A+B will therefore not be considered as reliable before adding new data to I, II and III. Incomplete combinations of the datasets, followed by a comparison of the results for all sets and combinations containing no data in common, is therefore a rigorous method that might allow the discovery and/or confirmation of clades.

Applications

In Fig. 2 we focus on a theoretical clade A+B of special interest. How was that clade selected? It was just by comparing phylogenetic results of different teams. Let us consider three research teams recovering the same clade A+B independently (Fig. 3) in their respective tree from simultaneous analysis. For the first team, the clade A+B is not reliable because it is present only in the analysis of the dataset I, but not in those of the datasets II and III. Then, the clade A+B is recovered by a second team (in the tree from the combined dataset IV+V+VI), and by a third team with still another dataset. This tends to indicate that the clade A+B is reliable, even if it was not repeated in the separate analyses of the first team, maybe because of stochastic errors. Figure 3 actually corresponds to a real situation. Dettaï & Lecointre and Miya *et al.* (2003b) both found a wider clade containing notothenioids. That clade was not repeated in Dettaï & Lecointre, the only one of the papers to use this confirmation protocol. Using the sequences of Dettaï & Lecointre, we explored the amount of confidence that ought to be given to this new group through the protocol explained in Figs 1 & 2. Four datasets have been analysed with that protocol: partial 12S-16S mitochondrial sequences, 28S nuclear ribosomal sequences, rhodopsin gene sequences (Chen *et al.* 2003, Dettaï & Lecointre), and the partial "Mixed Lineage Leukaemia-like" (called MLL hereafter) gene (Dettaï & Lecointre). The latter is a teleostean orthologue of a gene that, in humans, encodes a 4498 amino-acid long protein involved in leukaemogenesis (Caldas *et al.* 1998a, 1998b). Two additional datasets have been taken into account, but have not been analysed again. The resulting clades as published by the authors are presented in the table of repeatability (Table I): first, the tree obtained by Miya *et al.* (2003a, 2003b) based on the complete mitochondrial genome except for a few sequences (such as 12S and 16S rDNAs (so this dataset and that of Chen *et al.* (2003) do not overlap), and second, the tree from another fragment of MLL presented first in Venkatesh *et al.* (1999). Sequences used here have already been described in Chen *et al.* (2003), Miya *et al.* (2003a, 2003b), and Dettaï & Lecointre. Accession numbers of all sequences are given in Appendix 1. The 12S and 16S genes probably cannot be considered fully independent from one another because they are both constitutive of the mitochondrial ribosome and, therefore, potentially belong to the same selective unit. Consequently the 12S and 16S rDNAs were pooled into a single dataset. The two MLL datasets, as they are parts of a single gene, probably present the same problem. They could not be pooled, but it should be kept in mind when interpreting the results that they do not represent independent confirmations.

Alignments are the same as in Dettaï & Lecointre. They were performed by hand under BioEdit (Hall 1999). Alignments of new ribosomal sequence data were performed on the basis of alignments of Chen *et al.* (2003) based on secondary structure for the stem regions. The alignment of the loop regions in these datasets was based on several runs of Clustal X (Thompson *et al.* 1997) with default gap penalties and involving taxon order changes; it was then adjusted manually to avoid discontinuity of individual gaps as far as possible. Loops were conserved for

the analysis, but when gap length varied the gap regions were deleted. The rhodopsin sequences contain no gap, so alignment was not an issue. Alignment of MLL exons was easy to handle by eye because they encode a protein. Alignments and properties of each dataset (composition, saturation, etc.) are available upon request to the first author. A combined dataset was also created by concatenation of the sequences for each species. The size of each dataset, number of taxa, number of positions informative for parsimony, and all parameters resulting from most parsimonious tree searches are given Table II.

Analysis of repeatability

Table I shows clades resulting from phylogenetic analyses of separate and combined datasets, arranged so as to allow complementary comparisons. Complementary comparisons compare trees from independent datasets. That is all pairwise comparisons of datasets involving single datasets or combinations that do not contain data in common, but, put together, contain the four studied datasets (i.e. 28S against the combination of all datasets except 28S, or the combination of 12S-16S and 28S against the combination of MLL and rhodopsin). There are seven such complementary comparisons, four containing a single gene against the combination of the three others and three comparing combinations of the datasets by pairs. But the seven complementary sets are not the only ones taken into account. We also examined non-redundant, partially complementary subsets, such as the combination of 12S-16S and MLL when compared with 28S. In such cases, not all the data are used, but no data are shared, so information present in both parts of the set can be considered to be repeated. This is especially important as one subset can contain robust, but false information, and impose it on all the combined sets where it is present, possibly preventing a clade that the other datasets would support from ever appearing. We explored the presence of the widest common group to Miya *et al.* (2003b) and Dettaï & Lecointre, detected as in Fig. 3, then in complementary comparisons of trees of 104 taxa as shown in Fig. 2. We also focused on repeated patterns within that group.

Data analysis

Separate and simultaneous analyses have been conducted under the maximum parsimony method of inference of phylogeny. Under this criterion (hereafter called MP), heuristic searches were conducted with NONA v2 (Goloboff 1999a, 1999b) with Winclada (Nixon 1999-2002) as an interface, using TBR branch swapping. For a better exploration of tree space, the parsimony ratchet (hopper islands, Nixon 1999) was also used. The proportion of data to be weighted was set between 25% and 50% and the number of iterations progressively increased from 50 000 to 200 000 (option amb- poly=). This increase in iterations allowed us to make sure that the maximum number of MP trees had been reached. The number of trees was recorded after collapsing all unsupported nodes in all trees ("hard collapse"). Majority-rule consensus trees were used to summarize equally parsimonious trees obtained from each search. Heuristic searches run with PAUP4.0b10 (Swofford 1999) with the same parameters and a sufficient number of replicates yielded the same consensus trees. Bootstrap values were calculated for the combined data with PAUP4.0b10 on 1000 replicates repeated three times. This allowed confirmation that the results were very close in each run, i.e. that the number of replicates was high enough to obtain a good and repeatable approximation. See Fig. 4 for the values.

Results

Delineating the widest group common to the two studies

Comparing the combined tree in Miya *et al.* (2003b), based on more than 8000 mitochondrial characters, and Dettaï & Lecointre, based on 3525 nuclear and mitochondrial characters, we determined the broadest group with the same taxonomic content in both studies. Even though representative species differ, there are common genera or families. The single discrepancy in composition of the group is the position of *Trachinus draco*. In our trees it is always placed in this group, and never with the other trachinoids, but in Miya *et al.* (2003b) it groups with some of the other trachinoids (pinguipedids and ammodytids) within another larger group. Clade X is presented in the first line of Table I. It comprises percids, notothenioids, zoarcoids, cottoids, serranids, gasterosteids, scorpaenids, trachinids and triglids.

The data used in our simultaneous analysis differ from that in Dettaï & Lecointre only by the absence of the first fragment of MLL (Venkatesh *et al.* 1999, Dettaï & Lecointre), for which the taxonomic sampling was incomplete. Figure 4 illustrates the taxonomic samples used in these analyses. In this new simultaneous analysis, clade X is no longer present as a whole (see Fig. 4). The two studies differ in two non-robust nodes that make X paraphyletic: the clade that was the sister group of X in the Dettaï & Lecointre combined tree is still the closest to it, but is now inserted in it. However, as we do not establish reliability from the results of the combination of the data (Fig. 1), but from the repetition of the information across diverse datasets (Table I), we will assess the repeatability of clade X and its internal components across our four datasets and their various independent combinations. We also check these subgroups against the results for the almost complete mitochondrial genome of Miya *et al.* (2003b), and of the other fragment of MLL presented in Venkatesh *et al.* (1999) and Dettaï & Lecointre.

The sister group to our clade X is not easy to identify. In

Table I. Repeatability of clades across the datasets and datasets combinations (continued opposite).

Clades as defined by Chen (2001) and Dettai & Lecointre (in press)	Taxa and groups included	Maximum Parsimony 28S	All minus 28S	12S and 16S	All minus 12S-16S	Rhodopsin	All minus rhodopsin	Mll (Dettai & Lecointre in press)	All minus MLL	28S+ rhodospin	Rhodospsin +MLL	28S+ rhodospin	12S-16S+ MLL	12S-16S+ rhodospin	28S+ MLL
number of column		1	2	3	4	5	6	7	8	9	10	11	12	13	14
CLADE X:															
Is, K, P,Trachinus, Chelidonichthys, Scorpaena		E	?	no	X	no	E2	X	no	no	I	no	X	no	X
K, P or Pt		no	X	no	X	X	no	?	I	no	X	no	I	X	no
Is, K		I	no	no	no	no	X	?	no	X	no	no	no	no	no
Is, P		E	no	no	no	no	no	?	no	no	no	no	no	no	x
I	(Cottoidei,Zoarcoidei)	**X**	**X**	**X**	**X**	**X**	**X**	*I*	*X*	**X**	**X**	*x*	*I*	*X*	*I*
i1	Zoarcoidei	**X**	**X**	**X**	**X**	*X*	*I*	-	X	**X**	**X**	x	no	**X**	**X**
i2	Cottoidei	*E2*	*X*	*X*	*E*	*X*	*E*	*E*	*X*	*X*	*E*	x	no	*X*	*I*
Is	(Spinachia,Zoarcoidei)	no	no	no	no	no	X	X	no	no	no	no	no	no	no
Is	(I,Spinachia)	no	X	no	X	X	no	no	X	no	X	**x**	**X**	**X**	**X**
Isc	(Is, Chelidonichthys)	no	x	no	X	no	no	?	X	no	no	x	no	X	no
K	(k1,k2)	**X**	**X**	**X**	**X**	**X**	**X**	**x**	**X**	**X**	**X**	**x**	**X**	**X**	**X**
k1	Percidae	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**x**	**X**	**X**	**X**
k2	Notothenoidei	*I*	*X*	**X**	**X**	**X**	**X**	**X**	**X**	*X*	*E2*	*x*	*E*	*X*	*E*
k2, P		no	no	no	no	no	no	no	no	no	no	no	no	no	no
P	(Epinephelus, Holanthias, Rypticus, Pogonoperca, Serranus)	no	E	no	E	*E*	*I*	no	E	no	E	*E*	*E2*	*E*	*I*
Pt	P, Trachinus	no	no	no	X	no	no	?	no	no	X	no	X	no	no
G	(Ammodytes,Cheimarrichthys)	X	?	**X**	**X**	no	X	**X**	**X**	**X**	**X**	no	I	**X**	**X**
Gu	(G,Uranoscopus)	no	?	no	X	E	no	X	no	no	no	no	X	no	I
M	(Labrus,Scarus)	no	X	no	X	**X**	**X**	**X**	**X**	**X**	**X**	**x**	**X**	no	X
Mp	(Labrus,Scarus) Pomamadasys)	no	?	no	X	no	no	X	no	no	X	x	no	no	X
K, P, Trachinus, Scorpaena		no	?	no	E	E	no	?	I	no	no	x	no	X	no
Labrus, Scorp, Chelido, Trachinus		no	no	x	no	no	no	no	no	no	no	no	no	no	no
K, Scorpaena, Trachinus		no	no	no	no	no	no	?	I	no	no	I	no	no	E
Gu, Mp		no	?	no	X	no	no	x	no	no	no	no	no	no	X
G or Gu,Moronidae,Phycis		?	?	no	no	*E*	*I*	no	I	?	I	E	no	I	no
G, Mp, Lateolabrax, Phycis		no	?	no	no	no	no	?	no	no	x	no	no	no	no
Is, K, P or Pt, Mp, G or Gu, Lateolabrax, Phycis, Trachinus, Chelidonichthys, Scorpaena		E	?	no	I	no	X	?	no	no	I	no	no	no	?
K, P, Mp, G or Gu, Lateolabrax, Phycis, Trachinus, Scorpaena		no	?	no	no	no	no	?	no	no	no	?	no	no	no

Complementary datasets are put side to side in a single column. The presence of a clade in both parts of such a column is italicized if the clade is recovered with an inclusion or an escapee, in bold if it is recovered as is. X = clade present in the strict consensus tree, x = present in the majority rule consensus tree, E = clade present however a single taxon escapes, I = present but with insertion of another group inside, ? = not contradicted: group swamped into multifurcation, - = not enough taxa present to test the hypothesis. When a group is signalled with an inclusion or an escapee, details are given in the right hand column

both the Dettaï & Lecointre combined tree and ours, it is the group composed of Pinguipedidae (sandperches), Ammodytidae (sand lances), Uranoscopidae (stargazers), Moronidae (temperate basses) and *Phycis* (phycine hakes). The repeatability of this sister group relationship is very poor, even if the association of elements constitutive of these groups is much more frequent. *Phycis* is a gadiform that never groups with the other gadiforms using any of our data. Several independent amplifications and sequencings yielded the same result, and the Phycis sequence could not be checked against another phycid because none is available in the data banks yet, and the repeated Blast (Altschul *et al.* 1997) searches did not identify a taxon that could have produced contamination. This placement was unexpected, and this position should be considered as provisional until additional sequences are gathered. *Phycis* and the moronids are not present in the sampling of Miya *et al.* (2003b) sampling, but Pinguipedidae and Ammodytidae are within a group that is the sister group to clade X, forming a wider group that we do not recover in our trees. As our results or

Table II. (continued) Repeatability of clades across the datasets and datasets combinations.

Clades as defined by Chen (2001) and Dettai & Lecointre (in press)	Taxa & groups included	Maximum Parsimony Combination (Dettai & Lecointre in press)	Maximum Parsimony Combination of the 4 datasets	Maximum Parsimony Mll (Venkatesh *et al.* 1999)	Maximum Parsimony Miya *et al.* (2003)	Incoming and escaping taxa
number of column		15	16	17	18	
CLADE X:						
Is, K, P, Trachinus, Chelidonichthys, Scorpaena		X	no	*X*	*E*	1:Chelidonichthys; 6:Scorpaena,Trachinus; 10:Fistularia, Bothidae, Syngnathidae, Gobiidae, Mullus, Callionymus;18:Trachinus
K, P or Pt		no	I	?	?	8:Scorpaena, Uranoscopus;12:Trachinus;16:Scorpaena,Trachinus
Is, K		I	no	?	?	1:Holanthias; 15:Scorpaena, Serranus
Is, P		no	no	no	no	1: P only represented by Holanthias
I	(Cottoidei, Zoarcoidei)	X	*X*	*I*	*I*	12,14,18:Spinachia
i1	Zoarcoidei	X	**X**	-	**X**	6:Spinachia
i2	Cottoidei	X	**X**	-	**X**	1:Cyclopterus, Liparis;4,6:Liparis;14:Spinachia, Zoarcoidei
Is	(Spinachia, Zoarcoidei)	no	no	**X**	**X**	
Is	(I, Spinachia)	X	X	no	no	
Isc	(Is, Chelidonichthys)	X	X	-	I	18:Scorpaenidae
K	(k1,k2)	X	X	no	no	
k1	Percidae	X	**X**	**X**	**X**	
k2	Notothenoidei	X	**X**	**X**	**X**	1:Percidae;10:Pseudaphritis, Eleginops;12:Eleginops
k2, P		no	no	no	X	
P	(Epinephelus, Holanthias, Rypticus, Pogonoperca, Serranus)	E	*E*	*I*	*X*	E always Serranus; I always Scorpaena and Liparis
Pt	P, Trachinus	X	no	no	no	
G	(Ammodytes, Cheimarrichthys)	X	*X*	-	*I*	12:Uranoscopus; 18:Trachinus
Gu	(G, Uranoscopus)	X	**X**	**X**	-	5:Ammodytes; 14:Dicentrarchus
M	(Labrus,Scarus)	X	X	-	-	
Mp	((Labrus,Scarus) Pomamadasys	X	no	-	-	
K, P, Trachinus, Scorpaena		no	X	?	no	E always Scorpaena;8:Uranoscopus
Labrus, Scorp, Chelido, Trachinus		no	no	no	no	
K, Scorpaena, Trachinus		no	I	?	no	8:P,Uranoscopus;11:Trachinus;14:Scorpaena;16:P;
Gu, Mp		no	no	-	X	
G or Gu, Moronidae, Phycis		X	I	?	no	10:Mp;5,11:Phycis;6,8,13:Pomadasys;16:Dicentrarchus;
G, Mp, Lateolabrax, Phycis		no	X	-	-	
Is, K, P or Pt, Mp, G or Gu, Lateolabrax, Physcis, Trachinus, Chelidonichthys, Scorpaena		E	no	?	no	E always M;4:Dicentrarchus; 10:Fistularia,Bothidae,Syngnathidae,Gobiidae,Mullus,Callionymus
K, P, Mp, G or Gu, Lateolabrax, Phycis, Trachinus, Scorpaena		no	X	no	no	

Complementary datasets are put side to side in a single column. The presence of a clade in both parts of such a column is italicized if the clade is recovered with an inclusion or an escapee, in bold if it is recovered as is. X = clade present in the strict consensus tree, x = present in the majority rule consensus tree, E = clade present however a single taxon escapes, I = present but with insertion of another group inside, ? = not contradicted: group swamped into multifurcation, - = not enough taxa present to test the hypothesis. When a group is signalled with an inclusion or an escapee, details are given in the right hand column

results of Dettaï & Lecointre contradict the results derived from the (almost) complete mitochondrial genome, no conclusion can be drawn there. Additional nuclear data is needed to clarify this situation.

Repeatability within clade X

Previously identified groups

Some repeated components within the clade had already been identified in Chen *et al.* (2003), and Dettaï & Lecointre; these are indicated in Table I. Most of these clades do not include a great number of taxa, but several of them were unexpected with regard to previous morphological phylogenies.

From the present work the long sought group closest to Notothenioidei is consistent with the results of Chen *et al.* (2003). Percids are found in this place for all datasets and all possible combinations (see Table I & Fig. 4). This contradicts the results of Miya *et al.* (2003b), as they found

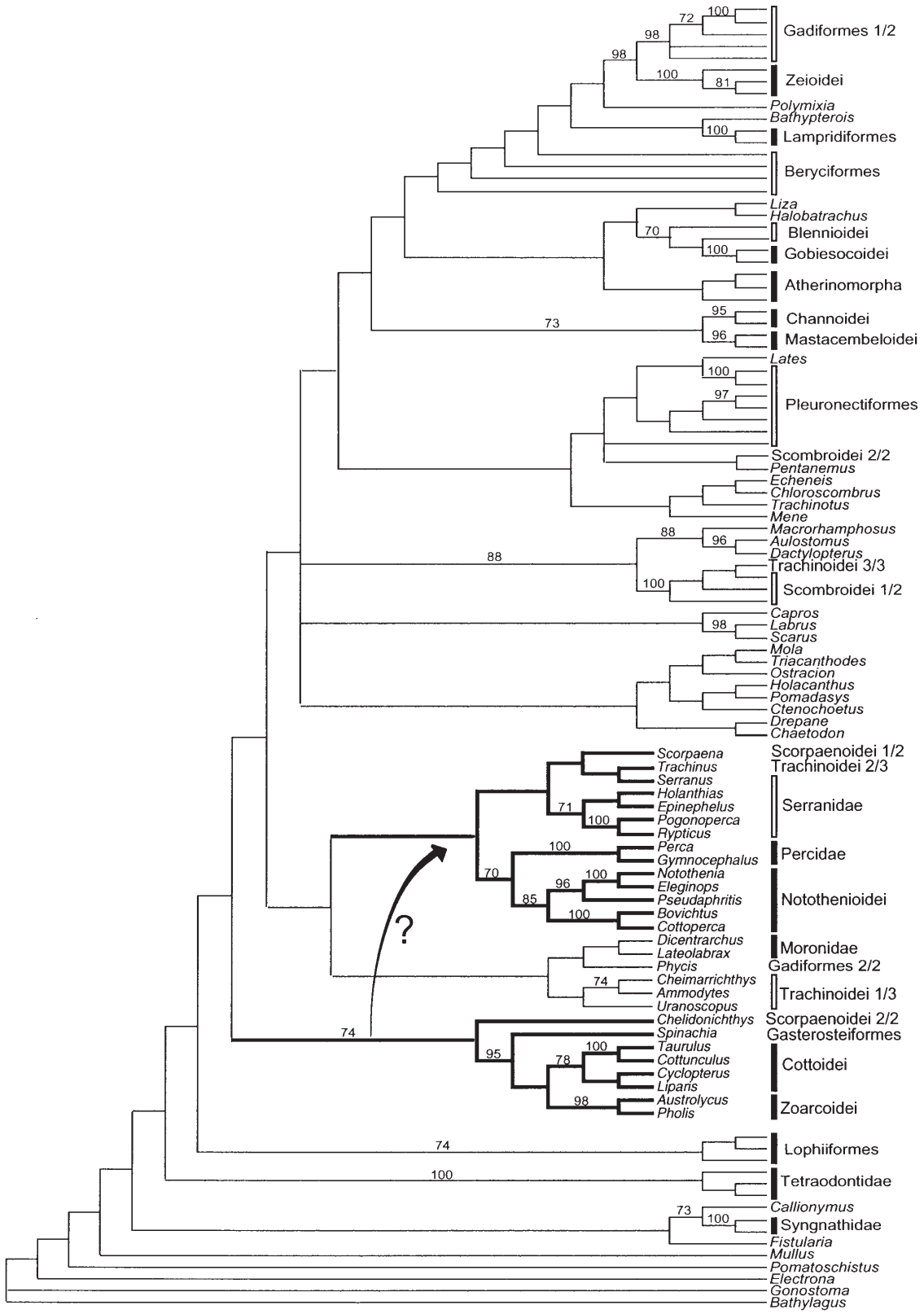

Fig. 4. Strict consensus of the 18 most parsimonious trees for the combined dataset. Length is 17 616 steps, CI = 17 and RI = 30. Numbers above nodes are bootstrap proportions calculated from 1000 iterations repeated several times superior to 70%. Monophyletic groups are indicated by black rectangles, non-monophyletic groups by white rectangles with the number of groups they are divided in indicated next to the name.

Table II. Size of each dataset, number of positions informative for parsimony, and parameters resulting from the parsimony searches.

	Nb of taxa	Analysed dataset length	Variable sites	Parsimony informative positions	Maximum parsimony (PAUP4) Number of trees after condense	Most parsimonious tree length	CI	RI
28S	95	821	373	241	423	1752	22	42
12S and 16S	104	740	545	458	36	7442	18	35
Rhodopsin	104	759	456	368	2	4916	20	36
Mll (new part)	92	527	385	318	2925	3253	20	35
28S + 12S-16S	104	1561	917	698	40	9411	18	32
28S + Rhodopsin	104	1580	829	609	812	6421	22	42
28S + MLL	104	1348	758	559	1216	5089	20	35
12S-16S + rhodopsin	104	1499	1001	826	4	12235	18	34
12S-16S + MLL	104	1267	930	776	3	11	17	32
Rhodopsin + MLL	104	1286	841	686	14	7919	17	30
All minus 28S	104	2026	1386	1144	8	15.674	18	32
All minus 12S-16S	104	2107	1214	927	7	9851	19	36
All minus rhodopsin	104	2088	1303	1017	54	13	19	35
All minus MLL	104	2320	1374	1067	18	14	17	30
Combined	104	2847	1759	1385	18	17616	17	30

serranids to be the closest, and percids as sister group to a Cottoidei-Scorpaenoidei-Zoarcoidei-Gasterosteidae clade (that is, the rest of clade X).

The clade I of Chen *et al.* (2003) that associates cottoids with zoarcids is found with the same consistency, except in a few cases where their sister group falls inside the clade. In Dettaï & Lecointre's work, that clade is included in a wider clade Is that also comprises *Spinachia*, the gasterosteid representative, as a sister group to clade I. In Miya *et al.* (2003a), *Gasterosteus,* also belonging to gasterosteids, is placed within clade I as the sister group to zoarcoids, a situation also found in three of our datasets: the combination of all data minus rhodopsin and the two MLL datasets when analysed on their own. However, the two MLL datasets cannot be considered as really independent from one another, as they originate from the same gene. *Chelidonichthys* (Triglidae) is associated as a sister group to the clade Is in five out of the fourteen datasets, but only in the ones containing the rhodopsin gene. It might well be imposed on the trees by the rhodopsin gene, and so cannot be accepted as reliable without additional data. In the tree from Miya *et al.* (2003a, 2003b), Scorpaenoidei (including triglids) are also the sister group to the zoarcoid-gasterosteid-cottoid branch. This relationship merits further study.

Classical groups

A very large quantity of data based on morphology generated many classifications and phylogenetic hypotheses, some of which are very well supported, while others are not. The molecular results do contradict some of these groups.

Serranids are rarely recovered as monophyletic. They are never recovered intact without insertions of other taxa (*Liparis* and *Scorpaena* most notably), and in most datasets *Serranus accraensis* does not group with the rest of serranids. However, neither does it group reliably with other taxa, so there is no reason for the moment to reconsider the monophyly of the family. Scorpaeniformes are never recovered, but even with morphological data the monophyly of the group was dubious at best, the suborbital stay, the only diagnostic character of the order, being non-homologous in Scorpaenoidei and Cottoidei (Nelson 1994, Imamura & Shinohara 1998). Scorpaenoidei are also very rarely recovered. The group is present only in trees derived from mitochondrial data: in trees of Miya *et al.* (2001, 2003a, 2003b) and in the trees from 12S-16S and 12S-16S + MLL combination, but not in trees obtained from the MLL dataset alone. Even in these cases, additional taxa always fall into this group, with the exception of the trees of Miya *et al*.

Conversely, some "expected" clades such as percids, notothenioids, zoarcids, cottoids are recovered for most datasets and most combinations. The monophyly of these groups can be considered to be confirmed by these molecular data.

New groups

Clade X was never recovered simultaneously by both parts of two complementary datasets. But it is recovered from several independent datasets, that is, both MLL fragments, the almost complete mitochondrial genome of Miya *et al.* (2003b), and 28S (with the exception of *Chelidonichthys* that escapes). It is also recovered from several combinations

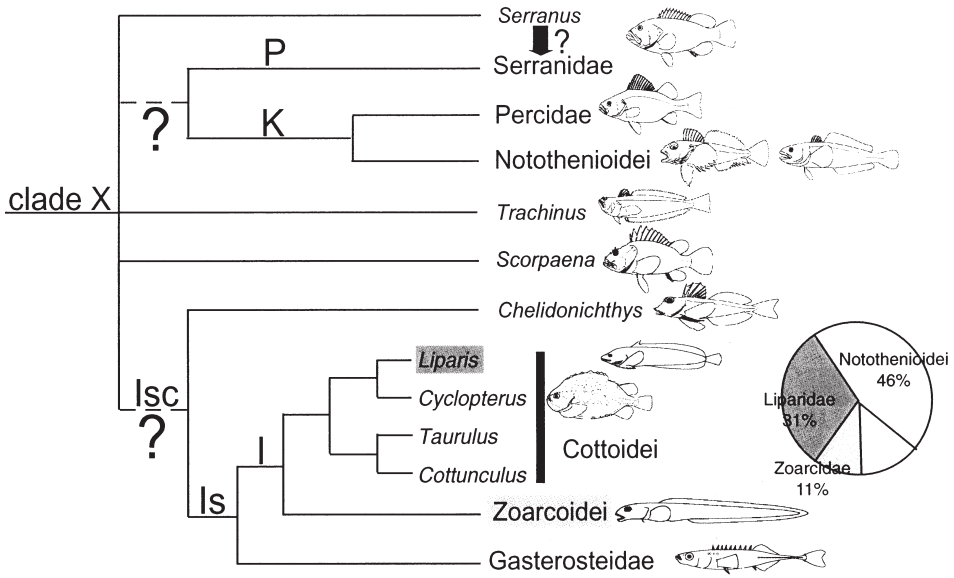

Fig. 5. Tree summarizing the taxonomic composition of clade X, with the proportion represented in the benthic fish fauna of the Antarctic shelf and upper slope (Eastman & Clarke 1998). Straight line clades within X are those considered as reliable because they are recurrently found from different datasets and research teams. Question marks indicate the groups that cannot yet be considered reliable, but are displayed here to indicate paths into further investigations. The designations of the clades according to Chen *et al.* (2003) and Dettai & Lecointre (in press) are given above the branches. P comprises serranids, K is the clade associating percids and notothenioids, I is the one comprising cottoids and zoarcoids, and Is and Isc the clades comprising respectively I plus gasterosteids and Is plus triglids. Grey squares indicate Antarctic taxa.

(see Table II). The most important fact to point out is that no alternative to clade X is repeated.

A partial clade X, comprising only serranids, percids, notothenioids, *Trachinus* and *Scorpaena* is found in the resulting tree of the rhodopsin data, leaving only one dataset, the 12S-16S, where only some components of the clade are recovered.

As for the arrangement of the groups within the clade, three main hypotheses can be considered, as there are three main subgroups (see lines 2–4 of Table I):

a) clade K (percids and notothenioids) with serranids,

b) clade Is with clade K, or

c) clade Is with serranids.

The first clade is the best supported, with seven datasets out of 14 supporting it; however, all these datasets except one (12S-16S+MLL) contain the rhodopsin data. This topology is the one found in our combined tree of the Fig. 4, but not in the combined tree of Dettaï & Lecointre. The other two alternatives are only found from two datasets each, one of which is a special case: for the 28S, the situation is ((Cottoidei, Zoarcoidei), Holanthias), (Percidae within paraphyletic Notothenioidei)), which could account for the second and third hypotheses. Still, additional support would be welcome to test these alternatives. Results of Miya *et al.* (2003b) do not bring much to the debate here, as they propose another alternative where clade K is split into two parts with percids associated with a Is, Scorpaenoidei group and notothenioids associated with serranids.

The positions of *Scorpaena* and *Trachinus* are too variable to be conclusive.

As a result, within X we consider as reliable the clades shown in Fig. 5: Zoarcoidei+Cottoidei (I), Gasterosteidae+I (Is), Notothenioidei+Percidae (K), and, to be confirmed, Serranidae+K and Triglidae + Is.

Discussion

Independence of multiple combinations

Not all pairwise combinations are independent of one another. Therefore, it would be erroneous to establish reliability from repeatability in such comparisons without discrimination. However, trees from pairwise combinations can be advantageously compared with trees from other independent pairwise combinations or trees from single datasets not already included in the combination without any loss of stringency of the test. In the same way, obviously trees from combinations containing all datasets minus one are to be compared with trees from the one that is absent. Incomplete combinations are used here only to check for decrease of stochasticity, not increase of repeatability.

The importance of a broad sampling must be stressed as well. Many reliable and unexpected groups have begun to emerge from the molecular results (i.e. gasterosteids with

cottoids and zoarcoids) because taxa they were made of have been gathered for the first time in a matrix. To a certain extent, all groupings proposed here are provisional: there is no way to know whether some acanthomorphs not sampled yet will fall into them, until all families are sampled. But we are confident that this study will be useful to highlight several points of interest for future research.

Are serranids monophyletic?

The monophyly of serranids is supported by four morphological shared specializations, three of them reductive and one innovative (Johnson 1983). However, none of the three first specializations is unique to serranids among percomorphs and not even their combination can characterize the group. As for the fourth character, the presence of the three opercular spines, it is much less common in other groups and, interestingly enough considering the composition of clade X, the only family actually presenting it in a state comparable to the one in serranids are the trachinids (Johnson 1983). As shown in Table II, *Trachinus draco* is found in a sister group position to serranids in three combinations and the combination of all the data. This relationship is not repeated in independent datasets and therefore cannot be considered as reliable, but it is nonetheless interesting and should be investigated with additional data.

The association of *Pogonoperca punctata* with the grammistine serranid *Rypticus saponaceus* in all datasets and combinations is consistent with Johnson's (1983) hypothesis placing it among Grammistini. The former hypothesis of Kendall (1976), giving *Pogonoperca* a basal position among serranids can be considered to be rejected by these molecular datasets. On the other hand, *Serranus* is not often recovered among serranids. In fact, the only time it is grouped with the other serranids is for the MLL and MLL + 28S datasets, but even in this case the group is recovered with intruders: *Liparis* and *Scorpaena*. We do not have enough data to question the monophyly of serranids, but it should be kept in mind that neither the morphological nor the molecular data give strong support to the group. The data of Miya *et al.* (2003b) does not allow any conclusion on this point, as their serranid representatives both belong to the genus *Epinephelus*.

Polyphyly of trachinoids

The delineation of trachinoids has been controversial for a long time (Gosline 1968, Pietsch 1989, Pietsch & Zabetian 1990, Johnson 1993, Nelson 1994, Mooi & Johnson 1997). Even considering only the taxa present in our sampling, the position of cheimarrichthyids, pinguipedids and chiasmodontids within trachinoids has been questioned by Johnson (1993) and Mooi & Johnson (1997) on morphological bases. Earlier molecular results (Chen *et al*. 2003, Dettaï & Lecointre) reliably placed *Kali*, a chiasmodontid, within scombroids, confirming the polyphyly of trachinoids *sensu* Pietsch & Zabetian (1990). These earlier papers also placed *Cheimarrichthys* repeatedly with *Ammodytes*, and *Uranoscopus* as a sister group to both (Chen *et al*. 2003, Dettaï & Lecointre). Although in the tree of Miya *et al*. (2003) *Parapercis* is also placed with *Ammodytes*, one cannot draw any conclusions as to the monophyly of a *Cheimarrichthys*-*Parapercis* group (Pinguipedidae *sensu lato*), as these have never been present together in any molecular dataset. As our results show, it cannot be ruled out that this part of the Trachinoidei might be the sister group of clade X along with Moronidae**.**

As detailed earlier, the only real incongruence in the composition of clade X with the data of Miya *et al.* (2003b) is on the position of *Trachinus draco*. With regard to what has been said previously about the existence of a morphological character supporting a serranid-trachinid relationship, the position of *Trachinus* in clade X makes some sense, but so does the placement of *Trachinus* with the other trachinoids in the conflicting tree generated from the mitochondrial data of Miya *et al.* (2003b). As in all cases of markedly contradictory data, additional information is necessary.

Rejection of classical sister groups for notothenioids

From our datasets and other molecular data, it can be concluded that a number of classical candidates for the sister group relationship to the Notothenioidei can be reliably rejected. Zoarcoidei (Anderson 1990) are closer to the Cottoidei. Trachinoidei (Gosline 1968) are polyphyletic, and the question of their relationships is complex because a complete sampling of all trachinoids for a molecular dataset does not yet exist.

Balushkin (2000) pointed out that a structure unique to pinguipedids and notothenioids, the antesupracleithral organ (a bilateral skin structure, situated in the dorso-caudal part of the branchial cavity, under the opercle, just in front of the supracleithrum), could be a synapomorphy of a group uniting both. Interestingly, *Cheimarrichthys*, sometimes included in the Pinguipedidae *sensu lato*, lacks this structure, and is not included in pinguipedids by Balushkin (2000). However, Miya *et al*. (2003b) found the pinguipedid *Parapercis* as the sister group of ammodytids, the same position we find here for *Cheimarrichthys*. Homoplasy of this character is confirmed by the fact that other authors do not agree on the interpretation of that organ. According to Eastman (personal communication 2003), the antesupracleithral organ is in fact the thymus, a lymphoid organ present in all the fishes. The observability of the thymus depends upon the preservation of the specimen. Eastman did observe the organ in *Cheimarrichthys.* It is therefore prudent to conclude that this character is not a reliable synapomorphy for a Pinguipedidae +

Notothenioidei group.

A clade that contains almost all Antarctic benthic fish fauna

By comparing results from different molecular studies and different datasets, we were able to extract a clade containing the notothenioids' closest relatives (Fig. 5), larger than just notothenioids and their sister group, and yet much better defined than the loose and too large Perciformes. Interestingly, that clade X contains 88% of the fish species recorded in the Antarctic continental shelf and upper slope (distributed in the three groups Zoarcoidei, Liparidae and Notothenioidei). Notothenioids are strictly Antarctic and sub-Antarctic, whereas zoarcoids and liparids are not restricted to Antarctic waters. The Liparidae have an anti-tropical distribution but are mostly known from the Northern Pacific, and the Zoarcoidei are more widely distributed but also occur primarily in the Northern Pacific. Within the clade X, Antarctic taxa indeed always have a non-Antarctic sister group. As can be seen in Fig. 5, the Zoarcoidei is not the sister group of the Notothenioidei but of the Cottoidei, which have a worldwide distribution, and the sister group to both are the coastal marine, brackish, or freshwater members of the northern hemisphere Gasterosteidae (Nelson 1994). Within cottoids, *Liparis* is the sister group of *Cyclopterus*, whose habitat is in the cooler regions of the seas and oceans of the northern hemisphere (Nelson 1994). And, of course, the group closest to notothenioids are percids, restricted to fresh water of the northern hemisphere, while the sister group to this clade are probably the serranids, inhabitants of tropical and temperate seas worldwide (Nelson 1994). Moreover, the three Antarctic benthic groups under focus all lack swimbladders, whereas their respective sister groups all have them. This supposes three independent adaptations to benthic life. In the present state of knowledge, and according to Fig. 5, we can infer that the Antarctic fish benthic community has had at least three distinct origins.

Acknowledgements

We thank Guido di Prisco, the EVOLANTA group of the Scientific Committee for Antarctic Research (SCAR) and the Comité Français pour la Recherche Antarctique for their support in communicating these new results. We also thank the French Polar Institute (Institut Paul E. Victor), for their logistic support for the scientific work on the field. We thank the referees, especially Professor J.T. Eastman, for their help in improving this paper.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25**, 3389–3402.

Anderson, M.E. 1984. *On the anatomy and phylogeny of the Zoarcidae.* PhD thesis, College of William & Mary, Williamsburg, VA, 253 pp. [Unpublished]

Anderson, M.E. 1990. The origin and evolution of the Antarctic Ichthyofauna. *In* Gon, O. & Heemstra, P.C., *eds*. *Fishes of the Southern Ocean*. Grahamstown, South Africa: J.B.L. Smith Institute of Ichthyology, 28–33.

Andriashev, A.P. 1964. A review of Antarctic fishes. *Issledovaniya fauny morei (Investigations of Marine Fauna),* **II**, 355–386.

Balushkin, A.V. 2000. Morphology, classification, and evolution of notothenioid fishes of the Southern Ocean (Notothenioidei, Perciformes). *Journal of Ichthyology,* **40**, S74–S109.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. & Wheeler, D.L. 2002. GenBank. *Nucleic Acids Research,* **30**, 17–20.

Berg, L.S. 1940. Classification of fishes, both recent and fossil. *Institute of Zoology of the Academy of Sciences of the USSR,* **5**, 87–517.

Berg, L.S. 1947. *Classification of fishes, both recent and fossil*. Ann Arbor, MI: J.W. Edwards Brothers, 87–517.

Bertin, L. & Arambourg, C. 1958. Super-Ordre des Téléostéens (Teleostei). *In* Grassé, P.P., *ed*. *Traité de Zoologie Tome XIII: Agnathes et Poissons* vol. 3. Paris: Masson et Cie, 2204–2500.

Caldas, C., Kim, M.H., Macgregor, A., Cain, D., Aparicio, S. & Wiedeman, L.M. 1998a. Isolation and characterization of a pufferfish MLL (Mixed lineage leukaemia like) gene (fMll) reveals evolutionnary conservation in vertebrate genes related to *Drosophila trithorax*. *Oncogene,* **16**, 3233–3241.

Caldas, C., So, C.W., Macgregor, A., Ford, A.M., McDonald, B., Chan, L.C. & Wiedemann, L.M. 1998b. Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene. *Gene,* **208**, 167–176.

Chen, W.J. 2001. *La répétitivité des clades comme critère de fiabilité : application à la phylogénie des Acanthomorpha (Teleostei) et des Notothenioidei (acanthomorphes antarctiques)*. PhD thesis, Paris VI University, Paris, 311 pp. [Unpublished]

Chen, W.J., Bonillo, C. & Lecointre, G. 2000. Taxonomic congruence as a tool to discover new clades in the acanthomorph (Teleostei) radiation. P. 369 in In *80th ASIH meeting* La Paz, Mexico, June 14–20 2000. Program book and abstracts, 400 pp. Where published and by whom?

Chen, W.J., Bonillo, C. & Lecointre, G. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution,* **26**, 262–288.

Dettai, A. & Lecointre, G. **Submitted.** New data confirm the recent clades obtained by multiple molecular phylogenies in the acanthomorph bush. *Journal of the Society of Integrative and Comparative Biology*.

Eastman, J.T. 1993. *Antarctic fish biology.* San Diego, CA: Academic Press, 322 pp.

Eastman, J.T. & Clarke, A. 1998. A comparison of adaptative radiations of Antarctic fishes with those of non-Antarctic fish. *In* Di Prisco, G., Pisano, E. & Clarke, A., *eds. Fishes of Antarctica: a biological overview.* Milan: Springer Italia, 3–26.

Froese, R. & Pauly, D. 2003. *FishBase*. www.fishbase.org/search/html (accessed in 2002).

Goloboff, P.A. 1999a. Analyzing large datasets in reasonable times: solutions for composite optima. *Cladistics,* **15**, 415–428.

Goloboff, P.A. 1999b. *NONA (NO NAME)* ver. 2. Tucuman, Argentina: Published by the author.

Gosline, W.A. 1968. The suborders of Perciform fishes. *Proceedings of the US National Museum,* **124**, 1–77.

Greenwood, P.H., Rosen, D.E., Weitzman, S.H. & Myers, G.S. 1966. Phyletic study of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History,* **131**, 339–456.

HALL, T.A. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* **41**, 95–98.

HASTINGS, P.A. 1993. Relationships of the fishes of the Perciform Suborder Notothenioidei. *In* MILLER, R.G., *ed*. *History and atlas of the fishes of the Antarctic Ocean*. Carson City, NV: Foresta Institute for Ocean and Mountain Studies, 99–107.

IMAMURA, H. & SHINOHARA, G. 1998. Scorpaeniform fish phylogeny: an overview. *Bulletin of the National Science Museum, Tokyo, Series A,* **24**, 185–212.

JOHNSON, G.D. 1983. *Niphon spinosus*: a primitive epinepheline serranid, with comments on the monophyly and intrarelationships of the Serranidae. *Copeia,* **3**, 777–787.

JOHNSON, G.D. 1993. Percomorph phylogeny: progress and problems. *Bulletin of Marine Sciences,* **52**, 3–28

JOHNSON, G.D. & PATTERSON, C. 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. *Bulletin of Marine Sciences,* **52**, 554–626.

KENDALL JR, A.W. 1976. Predorsal and associated bones in serranid and grammistid fishes. *Bulletin of Marine Sciences*, **26**, 585–592.

LECOINTRE, G. & DELEPORTE, P. 2000. Le principe du "total evidence" requiert l'exclusion de données trompeuses. *Biosystema,* **18**, 129–151.

LINDBERG, G.U. 1971. *Opredelitel'i kharakteristika semeistv ryb mirovoi fauny* [identification guide and characteristics of fish families of the world]. Leningrad, Russia: Nauka, 469 pp.

MAHNER, M. & BUNGE, M. 1997. *Foundations of biophilosophy*. Berlin: Springer, 423 pp.

MIYA, M., KAWAGUCHI, A. & NISHIDA, M. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial sequences. *Molecular Biology and Evolution,* **18**, 1993–2009.

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. & NISHIDA, M. 2003a. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution,* **26**, 121–138.

MIYA, M. & NISHIDA, M. 2003b. Major patterns of actinopterygian phylogenies: a new perspective based on >200 complete mitochondrial DNA sequences. Meeting Abstract, *Integrative & Comparative Biology*, **42**, 1280.

MIYAMOTO, M.M. & FITCH, W.M. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology*, **44**, 64–76.

MOOI, R.D. & JOHNSON, G.D. 1997. Dismantling the Trachinoidei: evidence of a scorpaenoid relationship for the Champsodontidae. *Ichthyological Research*, **44**, 143–176.

NELSON, J.S. 1994. *Fishes of the World*, 3rd ed. New York: John Wiley, 600 pp.

NIXON, K.C. 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics,* **15**, 407–414.

NIXON, K.C. 1999–2002. *WinClada*, ver. 1.0000. Ithaca, NY: Published by the author.

NIXON, K.C. & CARPENTER, J.M. 1996. On simultaneous analysis. *Cladistics,* **12**, 221–241.

NORMAN, J.R. 1937. Coast fishes. Part II: the Patagonian region. *Discovery Reports,* **16**, 1–150.

NORMAN, J.R. 1938. Coast Fishes. Part III: the Antarctic Zone. *Discovery Reports,* **18**, 1–104.

NORMAN, J.R. 1966. *A draft synopsis of the orders, families, and genera of recent fishes and fish-like vertebrates*. London: Trustees of the British Museum (Natural History), 649 pp

PIETSCH, T.W. 1989. Phylogenetic relationships of trachinoid fishes of the family Uranoscopidae. *Copeia,* **1989**, 253–303.

PIETSCH, T.W. & ZABETIAN, C.P. 1990. Osteology and interrelationships of the sand lances (Teleostei: Ammodytidae). *Copeia,* **1990**, 78–100.

REGAN, C.T. 1913. Classification of the Percoid fishes. *Annals of the Magazine of Natural History,* **8**, 111–145.

RIEPPEL, O. 2003. Popper and systematics. *Systematic Biology*, **52**, 259–271.

SOBER, E. 1988. Reconstructing the past: parsimony, evolution and inference. Cambridge, MA: The MIT Press, 265 pp.

SWOFFORD, D. 1999. *PAUP*, Phylogenetic Analysis Using Parsimony,* (* *and other methods*), version 4.0b10. Sunderland, MA: Sinauer Associates.

THOMPSON, J.D., GIBSON, T.J., PEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D.G. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* **25**, 4876–4882.

VENKATESH, B., NING, Y. & BRENNER, S. 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. *Proceedings of the National Academy of Sciences of the USA,* **96**, 10267–10271.

Appendix 1

Taxa included in this study and GenBank (Benson *et al.* 2002) accession numbers.
Classification following Nelson 1994 and listing order following the cladogramm proposed by Johnson & Patterson 1993. Indications are as follows:
Superorder: **order/suborder**: family, *Genus species* accession number for 28S, 12S, 16S, rhodopsin, MLL (Dettaï & Lecointre in press).

When a sequence is missing, the accession number is replaced by a -. For species indicated with ©, the sequenced specimen has been vouchered at the MNHN.

Outgroups: **Osmeriformes:** Bathylagidae, *Bathylagus euryops* AY141465-68, AY141325, AY141395, AY141255, -; **Stomiiformes:** Gonostomatidae, *Gonostoma atlanticum* AY141469-72(*G. bathyphilum*), D84033, D84049, AY141256(*G. bathyphilum*), -; **Chlorophthalmoidei**, Ipnopidae, *Bathypterois dubius* AY141473-76, AY141326, AY141396, AY141257, AY362219; **Myctophiformes**: Myctophidae, *Electrona antarctica* AY141477-80, AY141327, AY141397, AY141258, AY362201

Acanthomorpha : **Lampridiformes:** Lampridae*, Lampris immaculatus* AY141481-84, AY141328, AY141398, AY141259, -; Regalecidae, *Regalecus glesne* AY372729-30, AY368292, AY368296, AY368328, AY362266; **Polymixiiformes:** Polymixiidae, *Polymixia nobilis* AY372724-26, AF049730(*P. japonica*), AF049740(*P. japonica*), AY368320, AY362208;
Paracanthopterygii: **Gadiformes:** Gadidae,*Gadus morhua* AY141485-88, AY141329, AY141399, AF137211, -; *Merlangius merlangus* AY141489-92, AY141330, AY141400, AY141260, -; Phycidae, *Phycis blennioides* AY372733-36, AY368283, AY368306, AY368321, -; Moridae, *Mora moro* AY372739-42, AY368285, AY368307, AY368322, -; Macrouridae, *Coryphaenoides rupestris* AY372715-16, AY161233, AY368303, AY368319, -; *Trachyrincus murrayi* AY372708-10, AY368280, AY368301, AY368318, AY362289; **Lophiiformes:** Ceratiidae, *Ceratias holboelli* AY141505-08, AY141334, AY141404, AY141263, AY362270; Lophiidae, *Lophius piscatorius*© AY372751, AY368294, AY368305, AY368325, AY362274; Antennariidae, *Antennarius striatus*© AY372752-53, AY368287, AY368304, AY368324, AY362215; **Zeiformes:** Zeioidei, Zeidae, *Zeus faber* AY141493-96, AY141331, AY141401, Y14484, AY362287; *Zenopsis conchifer*© AY372748-50, AY368278, AY368300, AY368314, AY362286; Oreosomatidae, *Neocyttus helgae* AY141497-00, AY141332, AY141402, AY141261, AY362288; **Caproidei**: Caproidae, *Capros aper* AY141501-04, AY141333, AY141403, AY141262, AY362233
Beryciformes: **Trachichthyoidei**: Trachichthyidae, *Hoplostethus mediterraneus* AY141509-12, AY141335, AY141405, AY141264, AY362267; *Barbourisia sp*© -, AY368290, AF221881(*B. rufa*), AY368333, AY362264; **Berycoidei**: Berycidae, *Beryx splendens* AY141513-16, AY141336, AY141406, AY141265, AY362238; **Holocentroidei**: Holocentridae, *Myripristis botche* AY141517-20, AY141337, AY141407, U57539(*M. violacea*), AY362265; **Batrachoidiformes**; Batrachoididae, *Halobatrachus didactylus*© AY372743-44, AY368286, AY368308, AY368323, AY362246; ***Acanthopterygii***: **Percomorpha: Mugiloidei:** Mugilidae, *Liza sp.* AY141521-24, AY141338, AY141408, AY141266, AY362248; **Atherinomorpha: Bedotioidei**: Bedotiidae, *Bedotia geayi* AY141525-28, AY141339, AY141409, AY141267, AY362271; **Belonoidei**: Belonidae, *Belone belone* AY141529-32, AY141340, AY141410, AY141268, AY362273; **Cyprinodontoidei**: Poeciliidae, *Poecilia reticulata* AY141533-36, AY141342, AY141412, AY141269, AY362203; **Gasteresteriformes**: **Gasterosteoidei**: Gasterosteidae, *Spinachia spinachia* AY141585-88, AY141356, AY141426, AY141281, AY362261; **Syngnathoidei**: Aulostomidae, *Aulostomus chinensis* AY141577-80, AY141353, AY141423, AY141279, AY362226; Fistulariidae, *Fistularia petimba* AY372745, AY141355, AY141425, AY141324, -; Macroramphosidae, *Macroramphosus scolopax* AY141581-84, AY141354, AY141424, AY141280, AY362206; Syngnathidae, *Syngnathus typhle* -, AY368291, AF355009, AY368326, AY362211; *Hippocampus ramulosus*© -, AY368288, AY368310, AY368330, AY362216; **Synbranchiformes: Synbranchoidei**: Synbranchidae, *Monopterus albus* AY141565-68, AY141350, AY141420, AY141276, AY362252; **Mastacembeloidei**: Mastacembelidae, *Mastacembelus erythrotaenia* AY141561-64, AY141349, AY141419, AY141275, AY362249; **Dactylopteriformes**: Dactylopteridae, *Dactylopterus volitans* AY141589-92, AY141357, AY141427, AY141282, AY362243; **Scorpaeniformes: Scorpaenoidei**: Scorpaenidae, *Scorpaena onaria* AY141617-20, AY141364, AY141434, AY141288, AY362236; Triglidae, *Chelidonichthys lucerna* AY141609-12, AY141362, AY141432, AY141287, AY362284; **Cottoidei**: Cottidae, *Taurulus bubalis* AY141613-16, AY141363, AY141433, U97275, AY362217; Cyclopteridae, *Cyclopterus lumpus*© AY372737-38, AY368284, AY368299, AY368316, AY362218; *Liparis fabricii* -, -, -,AY368317, AY362235; Psychrolutidae, *Cottunculus gobio* AY372705-07, AY368279, AY368297, AY368315, AY362260; **Tetraodontiformes: Tetraodontoidei:** Tetraodontidae, *Lagocephalus laevigatus* AY141601-04, AY141360, AY141430, AY141285, AY362221; *Tetraodon nigroviridis* AJ270039-40-46, -, -, AJ293018, TN000000; *Takifugu rubripes* -,AJ421455, AJ421455, AF137214, AF036382, AF036382; Ostraciidae, *Ostracion sp*.© AY372722-23, AY368281, AY372754, AF137213, AY362207; Triacanthodidaedae, *Triacanthodes sp*.© -,AY368289, AY368311, AY368331, AY362258; Molidae, *Mola mola* AY141605-08, AY141361, AY141431, AY141286, AY362251; **Pleuronectiformes: Psettodoidei**: Psettodidae, *Psettodes belcheri*

AY372717-18, AY368282, AY368302, AF148143(*P. sp*), AY362259; **Pleuronectoidei**: Bothidae, *Arnoglossus imperialis* AY141593-96, AY141358, AY141428, AY141283, AY362228; *Bothus podas* AY372746-47, AF542221, AY157326, AY368313, AY362204; Paralichthyidae, *Syacium micrurum* -, -, -,AY368334, AY362262; Citharidae, *Citharus linguatula* AY372697-00, AF542220, AY157325, AY141323, AY362232; Soleidae, *Microchirus variegatus* AY141597-00, AY141359, AY141429, AY141284, AY362275; *Solea vulgaris* AY372727-28, AF542204, AF488442(*S. solea*), Y18672, -; **Perciformes: Percoidei**: Serranidae, *Serranus accraensis* AY141621-24, AY141365, AY141435, AY141289, AY362202; *Holanthias chrysostictus* AY141625-28, AY141366, AY141436, AY141290, AY362209; *Epinephelus aeneus* AY141629-32, AY141367, AY141437, AY141291, AY362227; *Rypticus saponaceus* -, AY368295, AY368309, AY368329, AY362257; *Pogonoperca punctata* AY372711-14, AY141368, AY141438, AY141292, AY362256; Centropomidae, *Lates calcarifer* AY141641-44, AY141371, AY141441, AY141294(2), -; Moronidae, Lateolabrax japonicus AY141633-36, AY141369, AY141439, AY141293, AY362253; *Dicentrarchus labrax* AY141637-40, AY141370, AY141440, Y18673, -; Percidae, *Perca fluviatilis* AY141645-48, AY141372, AY141442, AY141295, AY362279; *Gymnocephalus cernuus* AY141649-52, AY141373, AY141443, AY141296, AY362278; Chaetodontidae, *Chaetodon semilarvatus* AY372701-04, AF055592(*C. striatus*), AF055613(*C. striatus*), AY368312, AY362240; Drepanidae, *Drepane africana* AY141749-52, AF055595(*D. punctata*), AF055616(*D. punctata*), AY141321, AY362244; Pomacanthidae, *Holacanthus ciliaris* AY141753-56, AF055593, AF055614, AY141322, AY362214; Mullidae, *Mullus surmuletus* AY372719-21, AY368277, AF227680, Y18666, AY362231; Menidae, *Mene maculata* AY141729-32, AY141390, AY141460, AY141316, AY362250; Polynemidae, *Pentanemus quinquarius* AY141733-36, AY141391, AY141461, AY141317, AY362272; Haemulidae, *Pomadasys perotaei* -, AY368293, AY368298, -, AY362230; Carangidae, *Chloroscombrus chrysurus* AY141717-20, AY141387, AY141457, AY141313, AY362223; *Trachinotus ovatus* AY141721-24, AY141388, AY141458, AY141314, AY362263; Echeneidae, *Echeneis naucrates* AY141725-28, AY141389, AY141459, AY141315, AY362245; **Acanthuroidei**: Acanthuridae, *Ctenochaetus striatus* AY141745-48, AY141394, AY141464, AY141320, AY362242; **Labroidei**: Labridae, *Labrus bergylta* AY141737-40, AY141392, AY141462, AY141318, AY362222; Scaridae, *Scarus hoefleri* AY141741-44, AY141393, AY141463, AY141319, AY362212; **Zoarcoidei:** Zoarcidae, *Austrolycus depressiceps* AY141653-56,AY141374,AY141444,AY141297, -; Pholidae, *Pholis gunnellus* AY141657-60, AY141375, AY141445, AY141298, AY362285; **Notothenioidei**: Bovichtidae, *Bovichtus variegatus* AY141661-64, Z32702, Z32721, AY141299, AY362283; *Cottoperca gobio* AY141665-68, AY141376, AY141446, AY141300, -; *Pseudaphritis urvillii* AY141669-72, AY141377, AY141447, AY141301, -; Nototheniidae, *Notothenia coriiceps* AY141673-76, Z32712, Z32731, AY141302, AY362282; *Eleginops maclovinus* AY372731-32, AF145426, AF145411, AY141303, -; **Trachinoidei**: Trachinidae,*Trachinus draco* AY141681-84, AY141378, AY141448, AY141304, AY362277; Uranoscopidae, *Uranoscopus albesca* AY141685-88, AY141379, AY141449, AY141305, AY362239; Ammodytidae, *Ammodytes tobianus* AY141689-92, AY141380, AY141450, AY141306, AY362234; Cheimarrichthyidae, *Cheimarrichthys fosteri* AY141693-96, AY141381, AY141451, AY141307, AY362229; Chiasmodontidae, *Kali macrura* AY141697-00, AY141382, AY141452, AY141308, AY362224; **Blennioidei**: Blenniidae, *Parablennius gattorugine* AY141545-48, AY141345, AY141415, AY141271, AY362255; Tripterygiidae, *Forsterygion lapillum* AY141549-52, AY141346, AY141416, AY141272, AY362276; **Gobiesocoidei**: Gobiesocidae, *Lepadogaster lepadogaster* AY141553-56, AY141347, AY141417, AY141273, AY362247; *Apletodon dentatus* AY141557-60, AY141348, AY141418, AY141274, AY362213; **Callionymoidei**: Callionymidae, *Callionymus lyra* AY141541-44, AY141344, AY141414, AY141270, AY362225; **Gobioidei**: Gobiidae, *Pomatoschistus sp*. AY141537-40, AY141343, AY141413, X62405(*P. minutus*), -; **Scombroidei**: Sphyraenidae, *Sphyraena sphyraena* AY141713-16, AY141386, AY141456, AY141312, AY362254; Scombridae, *Scomber japonicus* AY141709-12, AY141385, AY141455, AY141311, AY362237; **Stromateoidei**: Stromateidae, *Pampus argenteus* AY141701-04, AY141383, AY141453, AY141309, AY362220; Centrolophidae, *Psenopsis anomala* AY141705-08, AY141384, AY141454, AY141310, AY362269; **Channoidei**: Channidae, *Channa striata* AY141569-72, AY141351, AY141421, AY141277, AY362241; **Anabantoidei:** Anabantidae, *Ctenopoma sp*. AY141573-76, AY141352, AY141422, AY141278, AY362210.