Structural Evolution of Otx Genes in Craniates

A. Germot,* 1 G. Lecointre,† J.-L. Plouhinec,* C. Le Mentec,* F. Girardot,* 2 and S. Mazan* 

*Equipe ATIPE de l’UPRES-A 8080 Développement et Evolution, Orsay, France; and †Laboratoire d’Ichtyologie et Service de Systématique moléculaire, Muséum National d’Histoire Naturelle, Paris, France

Using a degenerate PCR approach, we performed an exhaustive search of Otx genes in the reedfish Erpetoichthys calabaricus, the dogfish Scyliorhinus canicula, and the hagfish Myxine glutinosa. Three novel Otx genes were identified in each of these species, and their deduced protein sequences were determined over a large C-terminal fragment located immediately downstream of the homeodomain. Like their lamprey and osteichthyan counterparts, these nine genes display a tandem duplication of a 20–25-residue C-terminal domain, which appears to be a hallmark of all craniate Otx genes identified thus far, including the highly divergent Crx gene. Phylogenetic analyses show that, together with their osteichthyan counterparts, the dogfish and reedfish genes can be classified into three gnathostome orthology classes. Two of the three genes identified in each of these species belong to the Otx1 and Otx2 orthology classes previously characterized in osteichthysians. The third one unambiguously clusters with the Otx5/Otx5b genes recently characterized in Xenopus laevis, thus defining a novel orthology class. Our results also strongly suggest that the highly divergent Crx genes identified in humans, rodents, and oxen are the mammalian representatives of this third class. The hagfish genes display no clear relationships to the three gnathostome orthology classes, but one of them appears to be closely related to the LjOtxA gene, previously identified in Lampetra japonica. Taken together, these data support the hypothesis that the Otx multigene families characterized in craniates all derive from duplications of a single ancestral gene which occurred after the splitting of cephalochordates but prior to the gnathostome radiation. Using site-by-site sequence comparisons of the gnathostome Otx proteins, we also identified structural constraints selectively acting on each of the three gnathostome orthology classes. This suggests that specialized functions for each of these orthology classes were fixed in the gnathostome lineage prior to the splitting between osteichthysians and chordichthysians.

Introduction

Comparative analyses of an increasing number of chordate gene families have suggested that massive gene duplications have taken place during early vertebrate evolution, after the emergence of cephalochordates but before the splitting between chordichthysians and osteichthysians (defined as a group containing both actinopterygians and sarcopterygians, including tetrapods). Several authors have proposed that by releasing the constraints acting on duplicated genes, these genetic events may have facilitated the gain of new functions and substantially contributed to the morphological complexity displayed by vertebrates among chordates (Ohno 1970; reviewed in Holland et al. 1994). Orthodenticle-related homeobox genes, which have been identified with a high level of similarity over the homeodomain in representatives of all three major metazoan taxa, diploblasts (Müller, Yanze, and Schmid 1999; Smith et al. 1999), protostomes (Finkelstein et al. 1990; Bruce and Shankland 1998; Umesson, Watanabe, and Agata 1999), and deuterostomes, provide an example of such gene duplications. In deuterostomes, these genes, termed Otx genes, have been characterized in a wide range of species, including echinoderms (Gan et al. 1995), enteropneusts (Harada et al. 2000), ascidians (Wada et al. 1996), cephalochordates (Williams and Holland 1998), and a number of vertebrates. While they are present as single-copy genes in echinoderms, urochordates, and cephalochordates, small multigene families have been described in most vertebrates studied (references below). These gene duplications have been followed by functional diversifications of the different paralogs, as best illustrated in mice using targeted gene inactivation techniques. In this species, three paralogous Otx genes have been identified, Otx1, Otx2, and Crx. Otx1 and Otx2 display a dynamic expression pattern in the developing head and sense organs, with a characteristic expression domain at 10.5 days postcoitum (dpc) over the whole prosencephalon and mesencephalon (Simeone et al. 1992), while Crx is specifically transcribed in the retinal photoreceptors and the pinealocytes of the epiphysis (Furukawa, Morrow, and Cepko 1997). Each of these three genes is involved in some aspects of head development, but the associated mutant phenotypes substantially differ. Mouse Otx2, which, unlike Otx1 and Crx, is transcribed as early as 5.5 dpc in mouse embryos, is required during gastrulation for the initial induction of anterior neuroectoderm, a process which relies on cellular and molecular interactions with the adjacent visceral endoderm layer (Acampora et al. 1995; Rhinn et al. 1998). The resulting Otx2−/− phenotype, which is characterized in the neural tube by the specific deletion of anteriormost neuromes, shows some similarity to the deletion of the protocerebrum induced in Drosophila by an orthodenticle null mutation, thus suggesting that an ancestral role in brain specification may have been conserved between these phylogenetically distant species (Hirth and Reichert 1999). On a shorter evolutionary scale, among chordates, expression patterns of amphioxus and ascidian

1 Present address: Glycobiologie et Biotechnologie EA 1074, Faculté des Sciences, Université de Limoges, Limoges, France.

2 Present address: Génétique du Développement et Evolution, Institut Jacques Monod, Paris, France.

Key words: Otx, craniate phylogeny, gene duplication, dogfish, reedfish, hagfish.

Address for correspondence and reprints: Sylvie Mazan, Equipe ATIPE de l’UPRES-A 8080, Bât. 441, Université Paris-Sud, 91405 Orsay Cedex, France. E-mail: sylvie.mazan@ibaic.u-psud.fr.

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
Otx genes do support their involvement in anterior neuroectoderm specification through inductive interactions with mesendoderm (Wada et al. 1996; Williams and Holland 1996). Otx1-/- mice display a complex phenotype, characterized by a variety of defects, including pituitary hormone deficiencies and behavioral and morphological abnormalities (Acampora et al. 1996). Among the latter, the deletion of several gnathostome-specific characteristics, such as the lateral semicircular canal of the inner ear or the ciliary process, is of particular interest, since it suggests that the diversification of the Otx1 form may have been linked to their emergence during evolution (Mazan et al. 2000). Finally, in line with its expression pattern, Crx is essential for the differentiation of retinal photoreceptors and is also involved in the genetic control of circadian entrainment (Furukawa et al. 1999).

The molecular basis for these functional diversifications has been addressed by experiments consisting of the replacement of either Otx1 or Otx2 coding regions by the paralogous form in knock-in mice. These experiments have shown that differential transcriptional controls could largely, but not exclusively, account for the differences observed between Otx1-/- and Otx2-/- mutant phenotypes. When expressed under Otx2 endogenous regulatory sequences, Otx1 can completely restore the gastrulation defects observed in Otx2-/- mice. However, at later stages, the specification of the prosencephalon and the mesencephalon remains defective (Acampora et al. 1998; Suda et al. 1999). On the other hand, Otx2 can compensate for most Otx1-/- defects in knock-in mice except for the deletion of the lateral semicircular canal in all animals studied, and with lower penetrance for several eye defects, including the absence of ciliary canal in mice except for the deletion of the lateral semicircular canal of the inner ear or the ciliary process, is of particular interest, since it suggests that the diversification of the Otx1 form may have been linked to their emergence during evolution (Mazan et al. 2000). Finally, in line with its expression pattern, Crx is essential for the differentiation of retinal photoreceptors and is also involved in the genetic control of circadian entrainment (Furukawa et al. 1999).

A comparative analysis of vertebrate Otx genes in a wide range of species provides an alternative approach to address the molecular mechanisms underlying their functional diversification and possible links to gene duplication events. Most studies thus far have focused on osteichthyan species. Otx genes have been characterized in zebrafish (Li et al. 1994; Mori et al. 1994; Mercier et al. 1995), two anurans, Xenopus laevis (Pannese et al. 1995; Kablar et al. 1996) and Eleutherodactylus coqui (Fang and Elinson 1999), chick (Bally-Cuif et al. 1995), and several mammals (Simeone et al. 1992; 1993). In all of these species, genes clearly related to the mouse Otx1 and Otx2 paralogous forms have been unambiguously recognized, thus demonstrating the presence of two orthology classes among osteichthyan. Comparative analyses with amphioxus Otx have provided strong evidence that the corresponding duplication event took place prior to the splitting between actinopterygians and sarcopterygians but after the emergence of cephalochordates (Williams and Holland 1998). In contrast, Crx genes, which are divergent members of the gene family, have been found only in mammals (mouse, rats, humans, and oxen). Although these genes have aroused much interest in the past few years due to their involvement in human hereditary retinal degeneration, their origin is thus far completely unknown (Furukawa, Morrow, and Cepko 1997; Chen et al. 1999). In addition, other Otx genes have been isolated in Danio rerio (DrtOtx3; Mori et al. 1994; Mercier et al. 1995) and, recently, in X. laevis (Xlotx5 and Xot5b; Kuroda et al. 2000; Vignali et al. 2000). No orthological relationships have been described between these genes and any of their vertebrate counterparts, and the chronology of the corresponding duplication events is unclear. Finally, three Otx genes have also been identified in Lampetra japonica (LjOtxA and LjOtxB; Ueki et al. 1998) and Petromyzon marinus (PmOtx; Tomsa and Langeland 1999). Their relationship with their osteichthyan counterparts remains to be clarified.

To further characterize orthological relationships among vertebrate Otx genes and gain new insights into the structural evolution of craniate Otx proteins, we have started a characterization of these genes in two species which belong to phylogenetic groups of early emergence among craniates, the hagfish Myxine glutinosa and a chondrichthyan, the dogfish Scyllorhinus canicula. The redfish Erpetoichthys calabaricus, which belongs to the Cladista, a group of basal emergence among actinopterygians, has also been included in this analysis. We report here partial sequences of three Otx genes in each of these species. Their comparison with other vertebrate sequences provides new insights into orthological relationships among gnathostome Otx genes and into the structural constraints acting on their protein products.

Materials and Methods
Characterization of Otx Genes in E. calabaricus, S. canicula, and M. glutinosa

To identify Otx genes, a seminested degenerate PCR approach was taken, starting from genomic DNA and using the 5’- primers 5’-GTNTGGTTYAA-RAAYMG-3’ and 5’-GCNARTGYMGNCARCA-3’, which, respectively, encode the VWFKNR and AKCRQQ protein motifs, and the 3’ primer 5’-ARNA-CYTGRAAYTCC-3’, which encodes the WKFQVL protein motif. The cycling conditions were (1) denaturation, 95°C for 1 min; (2) annealing, 50°C for 1 min; and (3) elongation, 72°C for 2 min in the standard Taq DNA polymerase buffer. For each species, a heterogeneous population of amplified fragments of 550–700 bp in length were subcloned in linearized pTZ19R, and 24 independent recombinants were picked and sequenced. Sequences showing less than five nucleotide differences were inferred to derive from the same gene, with differences corresponding either to polymorphisms or to Taq DNA polymerase errors. The sequences corresponded to the consensus of a minimum of three independent clones.

Molecular Phylogenetic Analysis

Otx protein sequences were retrieved from the GenBank database. Their accession numbers are as follows: Branchiostoma floridae Otx, AF043740; P. marinus Otx, AF099746; L. japonica OtxA, AB012299; L. japonica OtxB, AB012300; D. rerio Otx1, D26172; D. rerio Otx2, D26173; D. rerio Otx3, D26174; L. laevis
were obtained from Kablar et al. (1996), *X. laevis* U77615; from proteins, allowing for the comparison of 31 sequences. This study were added to this alignment of chordate Otx (Philippe 1993). The nine new sequences determined in and *X. laevis* Otx, identified in an exhaustive alignment containing their cephalochordate and vertebrate counterparts (see Supplementary Material) immediately following the AKCRQQ motif, and the KXRX_{1-5}KXX motif (positions 16–23), which is also found in Amphi Otx. Except for one sequence of *M. glutinosa* (termed MgOtxD) which displays a divergent version (ILWNPA), they also contain the highly conserved (S/A)(I/L)WSPA motif (positions 111–116). This motif has been identified in a wide range of metazoan Otx proteins, including those encoded by the flour beetle *Tc-otd2* and the jellyfish *Pc-Otx* genes. Finally, sequences showing a similarity to the (D/E)CLDYK(D/E)(Q/P) motif (positions 365–376), called the Otx tail and present immediately upstream of the C-terminal W(K/R)FQVL motif in all deuterostome Otx proteins, can be unambiguously identified at the expected position in each case. The nine Otx genes thus identified were termed EcOtx1, EcOtx2, and EcOtx5 in *E. calabaricus*; ScOtx1, ScOtx2, and ScOtx5 in *S. canicula*; and *Mg-OtxA, MgOtxC*, and *MgOtxD* in *M. glutinosa*, according to their orthology relationships with their vertebrate counterparts (described below).

**Phylogenetic Analysis of Otx Genes in Craniates**

The partial Otx protein sequences determined in *E. calabaricus, S. canicula,* and *M. glutinosa* were included in an exhaustive alignment containing their cephalochordate and vertebrate counterparts (see Supplementary Material). All of the craniate sequences could be confidently aligned over extended regions of similarity (bold characters in the alignment), which were used for phylogenetic analyses. Outside of these regions, more variable intervening segments were excluded from the analysis. From the alignment of 380 amino acids, 148 informative sites for parsimony. Phylogenetic reconstructions were carried out using the NJ, MP, and ML methods (fig. 1). A single most-parsimonious tree (length = 479; retention index [RI] = 0.595; consistency index [CI] = 0.698) was obtained, and the length (L) of the ML tree was L = -2,615.9. Although the analyses were conducted unrooted, the resulting trees are shown rooted on *MgOtxC* and *MgOtxD* for convenience of presentation (fig. 1). All three analyses confirm the presence of the Otx1 and Otx2 orthology classes already characterized in osteichthynes. In addition to the Otx1- and Otx2-related genes identified in zebrafish, Xenopus, and mammals, each of these classes contains one of the genes newly characterized in *E. calabaricus* (*EcOtx1* and *EcOtx2*) and *S. canicula* (*ScOtx1* and *ScOtx2*). These two classes are always found to be monophyletic whatever the reconstruction method used. However, the clustering of the gnathostome Otx2 sequences is more highly supported (with BP values ranging from 75% to 97%) than the clustering of the Otx1-related paralogous genes (with BP reaching a maximum

**Results**

Identification of Three Otx Genes in *E. calabaricus, S. canicula,* and *M. glutinosa*

To identify Otx genes in *E. calabaricus, S. canicula,* and *M. glutinosa,* we used a PCR-based strategy, taking advantage of the high phylogenetic conservation of protein motifs located between residues 47 and 59 of the homeodomain (VWFKNR and AKCRQQ) and at the C-terminal ends of all vertebrate Otx proteins (W(K/R)FQVL). Degenerate amplification primers were designed to span these motifs, which delimit a large protein domain, ranging in size from 186 to 252 amino acids for Otx1 and Otx2, respectively. The corresponding coding region is contained within a single exon in mouse, as well as in amphioxus, suggesting that this genomic organization may be conserved among all vertebrates. All PCR amplifications were performed starting from genomic DNA.

In each studied species, this PCR strategy led to the amplification of a heterogeneous population of fragments, approximately 550–700 bp in length, which were subcloned and sequenced. Sequence comparisons of 24 subclones revealed the presence of three different fragments in the amplification products obtained from each of the three species under study, *E. calabaricus, S. canicula,* and *M. glutinosa.* These nine fragments unambiguously contain Otx coding sequences, as shown by their deduced amino acid sequences, which display a number of conserved motifs or residues previously identified in vertebrate Otx proteins. All share with their vertebrate counterparts a glutamine-rich stretch of variable length (Q_{1-5}, positions 1–5 in the alignment; see Supplementary Material) immediately following the AKCRQQ motif, and the KXRX_{1-5}KXX motif (positions 16–23), which is also found in Amphi Otx. Except for one sequence of *M. glutinosa* (termed MgOtxD) which displays a divergent version (ILWNPA), they also contain the highly conserved (S/A)(I/L)WSPA motif (positions 111–116). This motif has been identified in a wide range of metazoan Otx proteins, including those encoded by the flour beetle *Tc-otd2* and the jellyfish *Pc-Otx* genes. Finally, sequences showing a similarity to the (D/E)CLDYK(D/E)(Q/P) motif (positions 365–376), called the Otx tail and present immediately upstream of the C-terminal W(K/R)FQVL motif in all deuterostome Otx proteins, can be unambiguously identified at the expected position in each case. The nine Otx genes thus identified were termed EcOtx1, EcOtx2, and EcOtx5 in *E. calabaricus*; ScOtx1, ScOtx2, and ScOtx5 in *S. canicula*; and *Mg-OtxA, MgOtxC,* and *MgOtxD* in *M. glutinosa,* according to their orthology relationships with their vertebrate counterparts (described below).

**Phylogenetic Analysis of Otx Genes in Craniates**

The partial Otx protein sequences determined in *E. calabaricus, S. canicula,* and *M. glutinosa* were included in an exhaustive alignment containing their cephalochordate and vertebrate counterparts (see Supplementary Material). All of the craniate sequences could be confidently aligned over extended regions of similarity (bold characters in the alignment), which were used for phylogenetic analyses. Outside of these regions, more variable intervening segments were excluded from the analysis. From the alignment of 380 amino acids, 148 informative sites for parsimony. Phylogenetic reconstructions were carried out using the NJ, MP, and ML methods (fig. 1). A single most-parsimonious tree (length = 479; retention index [RI] = 0.595; consistency index [CI] = 0.698) was obtained, and the length (L) of the ML tree was L = -2,615.9. Although the analyses were conducted unrooted, the resulting trees are shown rooted on *MgOtxC* and *MgOtxD* for convenience of presentation (fig. 1). All three analyses confirm the presence of the Otx1 and Otx2 orthology classes already characterized in osteichthynes. In addition to the Otx1- and Otx2-related genes identified in zebrafish, Xenopus, and mammals, each of these classes contains one of the genes newly characterized in *E. calabaricus* (*EcOtx1* and *EcOtx2*) and *S. canicula* (*ScOtx1* and *ScOtx2*). These two classes are always found to be monophyletic whatever the reconstruction method used. However, the clustering of the gnathostome Otx2 sequences is more highly supported (with BP values ranging from 75% to 97%) than the clustering of the Otx1-related paralogous genes (with BP reaching a maximum...
of 67% for NJ). Inside the Otx2 class, the classical phylogeny of gnathostomes is recovered by the three reconstruction methods, except for the position of actinopterygians, which form a paraphyletic group in the MP tree (BP < 50%). In particular, the early emergence of S. canicula is always supported by high bootstrap values (93%, 61%, and 97% in NJ, MP, and ML analyses, respectively). The splitting of actinopterygians and sarcopterygians into two well-defined groups also receives strong support in NJ and ML analyses (BP > 62% and 77% for actinopterygians, 77% and 84% for sarcopterygians). Inside the Otx1 class, the branching orders differ substantially depending on the reconstruction method used. If the monophyly of mammals, as well as the close relationship of Otx1 and Otx2 sequences in X. laevis, are always recovered and supported by high bootstrap values, the relative position of the two actinopterygians (E. calabaricus and D. rerio) and the chondrichthyan (S. canicula) is not resolved.

In addition to the Otx1 and Otx2 classes, a third group, containing representatives of the major gnathostome taxa (chondrichthysans, actinopterygians, sarcopterygians including amphibians, and mammals) was found in all three tree reconstructions. This group contains the two closely related genes recently characterized in Xenopus (XIOtx5 and XIOtx5b) and the third gene identified in E. calabaricus and in S. canicula (thereafter named EcOtx5 and ScOtx5, respectively), together with the mammalian Crx genes. While the corresponding bootstrap values remain low in NJ and MP analyses, the existence of this third class receives good support in the ML analysis (BP = 91%). Inside this class, EcOtx5, XIOtx5, and XIOtx5b always form a monophyletic group, as do the mammalian Crx genes, but the relative positions of ScOtx5 and these two groups differ between the three reconstruction methods. This variability may be a consequence of the high divergence displayed by Crx sequences. However, Crx genes are always found in the same clade as Otx5 genes. The most parsimonious interpretation of these data is that Crx genes are orthologous to Otx5 genes but underwent an acceleration in their rate of evolution following the amphibian/amniote separation. In line with this hypothesis, we have been unable to find an Otx5 gene in the mouse or a Crx gene in Xenopus using the PCR strategy described here, whereas mouse Crx and Xenopus Otx5 gene fragments could be readily amplified by this method (data not shown).

As for the third gene identified in zebrafish (DrOtx3), it appears unrelated to the Otx1, Otx2, or Otx5/Crx classes in our tree reconstructions, always emerging before their separation. This result is incongruent with previously published studies which showed a clustering of DrOtx3 and DrOtx1 (Williams and Holland 1998).
However, the early emergence observed in our analyses for DrOtx3 is poorly supported. Furthermore, when only gnathostome sequences were included in the phylogenetic analyses, the zebrafish DrOtx1 and DrOtx3 sequences clustered together (data not shown). Together with the site-by-site analysis of the alignment in the more variable regions (see below), these results suggest that DrOtx3 may be a divergent member of the Otx1 class.

The lamprey and hagfish Otx sequences cannot be confidently assigned to any of the three classes identified in gnathostomes. Their relative branching orders appear variable depending on the reconstruction method used and most of the time are poorly resolved. A major exception lies in the clustering of LjOtxA with one of the hagfish genes (thereafter termed MgOtxA), which is observed in NJ, MP, and ML analyses. A close relationship between these two sequences is supported by high bootstrap values (96%, 95%, and 99% in NJ, ML, and MP analyses, respectively). In addition, a clustering of MgOtxD and PmOtx was obtained in the three phylogenetic reconstructions with bootstrap values above 50%. However, because of the high degree of divergence of these two sequences, we cannot exclude that this result may be due to a long-branch attraction artifact.

Phylogenetic Analysis of Amphioxus and Craniate Otx Sequences

In an attempt to root the craniate Otx phylogenetic tree, amphioxus Otx sequence was added for a new phylogenetic analysis (fig. 2). The portions of the alignment which were taken into account were the same as those previously used. Among the 150 sites used in this analysis, 136 corresponded to variable sites and 102 were informative for parsimony. As in the previous analysis, phylogenetic reconstructions were carried out using the NJ, MP, and ML methods. Ten equiparisonious trees were found (length = 517; RI = 0.684; CI = 0.594), and the length of the ML tree was \( L = -2,790.86 \).

The three gnathostome Otx classes and the cyclostome OtxA class were recovered, with slightly less support than in the previous analyses. The ML analysis provided the strongest support for the monophyly of the Otx2 and Otx5/Crx classes (96% and 92%, respectively), but weak support concerning Otx1-related genes (41%). In the MP analysis, the three classes were recovered in the 10 MP trees found, albeit with relatively low bootstrap values (38%, 63%, and 42%, respectively, for the Otx1, Otx2, and Otx5/Crx classes). The NJ analysis also confirmed the presence of the three classes, with relatively low bootstrap values for the Otx1 and Otx5/Crx classes (59% and 33%) and a higher one for Otx2 (92%).

The close relationship observed between LjOtxA and MgOtxA in the previous analysis was recovered with high bootstrap values in all three reconstruction methods. No clear conclusion could be reached concerning the relative order of emergence of the three gnathostome orthology classes, the cyclostome OtxA class, and the other lamprey and hagfish genes (PmOtx, Lj-OtxB, MgOtxC, and MgOtxD). It appeared variable depending on the reconstruction method used (data not shown), and none of the groupings appeared to be supported by significant bootstrap values (fig. 2).

Structural Evolution of the Otx1, Otx2, and Otx5/Crx Orthology Classes in Gnathostomes

To independently assess the existence of the three orthology classes identified in the phylogenetic analysis and gain deeper insight into their structural evolution, we performed a systematic search of motifs or residues specifically maintained in each of them. This analysis was carried out for the portions of the alignment which were included in the phylogenetic analysis and extended to the more variable intervening regions (see alignment). In the latter, no strongly supported alignment can be proposed between all gnathostome Otx proteins, but significant similarities can be identified within each of the four classes, respectively, containing Otx1/Otx4, Otx2/Otx2b, Otx5/Otx5b, and Crx sequences. In particular, the highly variable regions (positions 59–105 and 121–193) located immediately upstream and downstream of the WSP motif show an extensive divergence between Otx1, Otx2, Otx5, and Crx proteins but can be unambiguously aligned within each of these four classes. These close relationships strongly support the assignment of Sc/EcOtx1, Sc/EcOtx2, and Sc/EcOtx5 to the Otx1, Otx2, and Otx5 orthology classes, respectively. Most of the repetitive stretches of serine and histidine (positions 62–75, 257–263, 290–299, and 307–317), previously identified as hallmarks of osteichthyan Otx1-related genes, are also conserved in EcOtx1 and ScOtx1 but absent from all other craniate sequences, suggesting that they correspond to synapomorphies of the Otx1 orthology class. The only exception is the histidine stretch located at positions 290–299, which appears to be restricted to a single additional H in ScOtx1. From position 257 to position 299, ScOtx1 sequence also provides new clues for understanding the structural evolution of Otx1-related genes, since it displays a 7-amino-acid motif (positions 270–276) which has been deleted in all osteichthyan Otx1 proteins but is present in all other craniate sequences. While the relationship between osteichthyan Otx1 sequences and other vertebrate sequences remained unclear over this domain, the identification of this ancestral characteristic makes it possible to confidently align them with their craniate counterparts.

In gnathostomes, the characterization of Otx1-, Otx2-, and Otx5-related coding sequences in a wide range of species, including in all cases a chondrichthyan, at least one actinopterygian, and one osteichthyan, makes it possible to identify residues which have been selectively conserved within each of these three classes. A total of 15, 10, and 14 such amino acids, respectively, selectively maintained among Otx1, Otx2, and Otx5 protein sequences, could be identified. These residues are shaded in the alignment (Supplementary Material). They are present not only in the most variable regions of the proteins, which were excluded from the phylogenetic analysis, but also within the protein domains.
Using these distinctive features as a criterion, we studied the assignment of the zebrafish Otx3 and the mammalian Crx genes to the Otx1, Otx2, and Otx5 classes identified in gnathostomes. Over the protein domain studied, DrOtx3 displays 10/15 (positions 41, 43, 61, 62, 123, 219, 257, 262, 263, and 308) and 1/10 residues (position 284) specific for Otx1- and Otx2-related genes, respectively, but no residue specific for the Otx5 class. This suggests that it may be a divergent member of the Otx1 orthology class. In line with this conclusion, it ex-
hibits the serine and histidine stretches which are also characteristic for this class (positions 62–74, 257–266, 290–294, and 307–313). As for the four mammalian Crx genes, their deduced amino acid sequences share 9 selectively conserved residues out of 14 with Otx5 sequences but appear unrelated to Otx1 or Otx2 forms over the residues specific for these classes. This result thus supports their orthological relationship with Otx5 genes, as already pointed out by the phylogenetic analysis.

The Otx sequences characterized in lamprey and in hagfish show no clear relationship with either the gnathostome Otx1, Otx2, or Otx5 form over the residues specific for each class. In contrast, L. japonica and M. glutinosa OtxA proteins share a total of eight residues which are not found in any other craniate Otx form. While the ancestral state cannot be determined in most cases, the G (position 358) which corresponds in both LjOtxA and MgOtxA to the N shared by the cephalochordate and all other craniate Otx sequences (except Crx forms, which display a T) is likely to correspond to a derived character.

Otx Tail Duplication in Craniate Otx Proteins

Phylogenetic reconstructions taking the whole protein sequence into account have provided strong evidence that the duplication events which have given rise to the lamprey Otx genes and the gnathostome Otx1 and Otx2 orthology classes took place after the splitting between the cephalochordate and vertebrate lineages. In line with this conclusion, it has been shown that these genes share sequence features which are not observed in echinoderms, ascidians, or cephalochordates. In particular, all of them display an imperfect tandem duplication of an 18–25-amino-acid motif containing the Otx tail (domains A and B in the alignment; see Supplementary Material). In contrast, a single copy is present in echinoderms, ascidians, and amphioxes (Williams and Holland 1998).

Like their orthologs in sarcopterygians and in zebrafish, the Otx1 and Otx2 coding sequences identified in E. calabaricus and in S. canicula contain two copies of the C-terminal domain. To address the chronology of the duplication events that have generated the Otx5/Crx orthology class identified in gnathostomes and the multigene family characterized in hagfishes, we examined each of the sequences reported in this study for the presence of this C-terminal repeat. Both domains A and B are unambiguously present at the C-terminal ends of the four Otx5 proteins identified in X. laevis, E. calabaricus, and S. canicula (positions 329–356 and 357–380). However, the presence of this tandem duplication has not been recognized in the amino acid sequence encoded by mammalian Crx genes (Furukawa, Morrow, and Cepko 1997). Since they cluster with Otx5 genes in our phylogenetic reconstructions, we searched for the presence of a duplicated C-terminal domain in Crx sequences. Despite a 4-amino-acid deletion in the N-terminal part of domain A, all four mammalian Crx sequences contain the DSLEFKDPTGTWK motif, which can be unambiguously aligned with the C-terminal domain (positions 337–356). Immediately downstream (positions 357–380), domain B can also be unambiguously recognized, although its N-terminal consensus in Otx1, Otx2, and Otx5 sequences (LNF, positions 357–359) is replaced by the motif FTY in Crx proteins. Although degenerate, the C-terminal repeat is thus present in Crx amino acid sequences. In hagfish, the three sequences identified also contain both domains A and B at their C-terminal ends (positions 329–356 and 357–380), with several single-amino-acid insertions or deletions.

Discussion

Orthological Relationships Among Craniate Otx Genes

We report here the identification and partial sequencing of three Otx genes in a polypterid, the reedfish E. calabaricus; a chondrichthyan, the dogfish S. canicula; and an agnathan, the hagfish M. glutinosa. Concerning the orthological relationships among gnathostome Otx genes, two strongly supported conclusions can be drawn from the phylogenetic analyses of these sequences. First, Otx1- and Otx2-related genes are present in all gnathostomes studied thus far, including the chondrichthyan S. canicula. Among extant vertebrates, a strict correlation is thus observed between the appearance of some gnathostome-specific characteristics which appear to be deleted in Otx1− mice (lateral semicircular canal of the inner ear, ciliary process) and the emergence of Otx1-related genes. This result supports the hypothesis that the diversification of gnathostome Otx genes is linked to the rise of these innovations during evolution (Mazan et al. 2000). Second, we show the presence of a novel orthology class, termed Otx5, in gnathostomes: this class unambiguously contains the Otx5/Otx5b genes recently identified in Xenopus, together with the third gene identified in E. calabaricus and S. canicula. In addition, our results also indicate that the highly divergent Crx genes, characterized in humans, rodents, and oxen, may be the mammalian representatives of this third group. In line with this hypothesis, Xenopus Otx5/Otx5b genes and rodent Crx genes share highly specific expression domains in the developing eye and epiphysis (Furukawa, Morrow, and Cepko 1997; Vignali et al. 2000). Such expression patterns substantially differ from the broad expression domains, spanning the whole prosencephalon and diencephalon displayed by all gnathostome Otx1 and Otx2 genes, thus supporting the relationship proposed between Otx5 and Crx genes. Finally, in the hagfish M. glutinosa, as in gnathostomes, Otx genes form a small multigene family, comprising at least three paralogs. Like their lamprey counterparts, these genes show no clear relationship to the three orthology classes identified in gnathostomes. However, one of them is closely related to the LjOtxA gene previously isolated in L. japonica. This close relationship, which was highly supported by all our phylogenetic analyses, may simply reflect the clustering of both species in the same phylogenetic group. In line with this hypothesis, rRNA comparisons support the grouping of lampreys and hagfishes in the same phylogenetic group, the cyclostomes, which
is a sister group of the gnathostomes (Stock and Whitt 1992; Mallatt and Sullivan 1998). However, other analyses point to the basal emergence of hagfishes, considered a sister group of the vertebrates (lampreys and gnathostomes) (Janvier 1996; Rasmussen, Janke, Arnason 1998; Delarbre et al. 2000). Since no OtxA-related gene can be recognized among gnathostomes, this phylogeny would imply that OtxA-related genes either have been lost or have undergone a particularly high divergence rate in the gnathostome lineage after the splitting of lampreys.

**Gene Duplications in the Otx Gene Family**

The basal emergence of the chondrichthyans has recently been challenged on the basis of mitochondrial DNA analyses (Rasmussen and Arnason 1999). However, both the phylogenies obtained for Otx2 and Otx5 genes and the site-by-site analysis of gnathostome Otx1 sequences strongly support the basal position of the chondrichthyans as a sister group of osteichthyan. In this phylogeny, the identification of S. canicula Otx genes belonging to the three orthology classes characterized in osteichthyans indicates that the corresponding duplication events took place prior to the splitting between chondrichthyans and osteichthyan, more than 420 MYA (Benton 1990). The tandem repeat of the C-terminal motif observed in Otx1- and Otx2-related proteins, as well as in the three Otx proteins identified in L. japonica and P. marinus, has provided a strong argument that the corresponding genes all derive from duplications of a single ancestral gene which occurred after the splitting of cephalochordates (Ueki et al. 1998; Williams and Holland 1998; Tomsa and Langeland 1999). This conclusion can now be extended to the gnathostome Otx5/Crx genes and the three copies identified in hagfish, suggesting that the presence of this C-terminal repeat is a synapomorphy of all craniate Otx proteins. In contrast, the chronology of the duplication events that gave rise to the three orthology classes characterized in gnathostomes relative to the emergence of lampreys and hagfishes could not be resolved in our analyses. Independent duplication events provide the simplest hypothesis to account for the absence of sequence relationships between the lamprey or hagfish genes and the gnathostome Otx1, Otx2, and Otx5/Crx orthology classes. However, the possibility that gene duplications occurred prior to the splitting of lampreys and hagfishes cannot be formally ruled out. A relatively short time interval between such genetic events and the speciation events which generated these lineages could actually make ontological relationships with the gnathostome Otx genes difficult to detect.

These results are very similar to those obtained in a number of other genetic systems in two respects. First, they show the presence of unambiguous orthology classes in all jawed vertebrates, including chondrichthyans. Similar conclusions have been reached in a number of other molecular studies also including chondrichthyan sequences (Sidow 1992; Duguay et al. 1995; Kandil et al. 1996; Li, Keith, and Evans 1996; Hahn et al. 1997; Flajnik et al. 1999). Second, in hagfish, as in lamprey, they provide an additional example of a multigene family resulting from duplications of a single ancestral copy in the craniate lineage. An absence of clear ontological relationships between the paralogous genes identified in these species and their gnathostome counterparts has also been a recurrent conclusion for the different studied systems (Sidow 1992; Li, Keith, and Evans 1996; Sharman and Holland 1998; Kuraku et al. 1999; Suga et al. 1999). Taken together, these results support the hypothesis that the duplication of Otx genes may have been part of massive amplification events which took place after the emergence of cephalochordates and prior to the splitting between chondrichthyans and osteichthyan (Williams and Holland 1998). However, a more precise chronology of such genetic events remains elusive.

**Functional Evolution of Gnathostome Otx Genes**

Several lines of evidence indicate that for some of the complex functions fulfilled by Otx genes in gnathostomes, paralogous Otx proteins may share largely equivalent biochemical properties. For instance, all the gastrulation defects observed in Otx2−/− mouse embryos, including the initial induction of anterior neuroectoderm, can be compensated when the mouse Otx2 coding region is replaced by its Otx1 counterpart in knock-in experiments (Acampora et al. 1998; Suda et al. 1999). Likewise, the replacement of Otx1 by Otx2 restores most of the brain and sense organ abnormalities observed in Otx1−/− mice (Acampora et al. 1999). In Xenopus, the phenotypes induced by overexpressions of XIOtx2 and XIOtx5/Otx5b are also very similar, suggesting that both paralogs are involved in the early specification of anterior regions (Andreazzoli, Pannese, and Boncinelli 1997; Kuroda et al. 2000; Vignali et al. 2000). On the other hand, the identification of amino acid differences between Otx1, Otx2, and Otx5 proteins, which appear to be selectively maintained among a wide range of gnathostomes, provides strong evidence that different structural constraints act on each orthology class. Whatever the precise chronology of Otx genes duplication events in craniates, these structural constraints were fixed after the cyclostome/gnathostome splitting but prior to their radiation. This suggests that functions which are unique to each orthology class were also fixed in the gnathostome lineage, prior to the chondrichthyan/osteichthyan splitting. Results of knock-in experiments and comparisons of the expression patterns of gnathostome Otx genes are consistent with this hypothesis. The role played by mouse Otx1 in the lateral semicircular canal formation could provide an example of such specializations, since it cannot be compensated for by the paralogous form Otx2 in knock-in experiments. Concerning Otx2, it is not currently known whether the jaw defects observed in Otx2−/− mutant mice (Matsuo et al. 1995) can be compensated by paralogous genes. In contrast, the replacement of Otx2 coding sequence by the paralogous Otx1 form in mice results in a defective prosencephalon specification during neurulation, raising the possibility that this role may be specific to the Otx2
orthology class. As for the Otx5/Crx orthology class, it may have been recruited for specific roles in photoreceptor development, as suggested by the specific transcription pattern shared by amphibian Otx5 genes in the developing eye and epiphysis and the analysis of Crx−/− mice. The analysis of ScOtx5 expression should help to test this hypothesis.

Taken together, these results suggest that while some roles of Otx genes may be fulfilled by different combinations of paralogous genes depending on the species considered, each gnathostome orthology class has also been recruited for specific functions. Residues corresponding to selectively maintained differences between the three orthology classes may be important in determining the specificity of the molecular interactions which control these processes. The identification of such interactions, together with detailed analyses of the structure/function relationships of gnathostome Otx genes, will be required to elucidate the genetic and molecular mechanisms underlying their functional diversification.

**Supplementary Material**

Nucleotide sequences encoding the protein fragments reported in this study for the reedfish EcOtx1, EcOtx2, and EcOtx5, the dogfish ScOtx1, ScOtx2, and ScOtx5, and the hagfish MgOtxA, MgOtxC, and MgOtxD have been deposited in the GenBank database under accession numbers AF321406, AF321407, AF321408, AF321409, AF321410, AF321411, AF321412, AF321413, and AF321414, respectively. The alignment of the craniate and amphioxus Otx sequences are shown in supplementary material on the SMBE web site.

**Acknowledgments**

This work was supported by Fondation pour la Recherche Médicale and Société de Secours des Amis des Sciences postdoctoral fellowships to A.G. and funded by the Centre National de la Recherche Scientifique, the Université Paris-Sud, and the Fondation pour la Recherche Médicale. We thank Hervé Philippe for advice on phylogenetic reconstructions, Noëlle Narroad for her help in typing the manuscript, Lenny Rabinow for his careful reading of the manuscript, and Monique Pradel for her excellent technical assistance.

**LITERATURE CITED**


WILLIAM JEFFERY, reviewing editor

Accepted May 11, 2001