We report the characterization of three Emx genes in a chondrichthyan, the dogfish Scyliorhinus canicula. Comparisons of these genes with their osteichthyan counterparts indicate that the gnathostome Emx genes belong to three distinct orthology classes, each containing one of the dogfish genes and either the tetrapod Emx1 genes (Emx1 class), the osteichthyan Emx2 genes (Emx2 class) or the zebrafish Emx1 gene (Emx3 class). While the three classes could be retrieved from the pufferfish genome data, no indication of an Emx3-related gene in tetrapods could be found in the databases, suggesting that this class may have been lost in this taxon. Expression pattern comparisons of the three dogfish Emx genes and their osteichthyan counterparts indicate that not only telencephalic, but also diencephalic Emx expression territories are highly conserved among gnathostomes. In particular, all gnathostomes share an early, dynamic phase of Emx expression, spanning presumptive dorsal diencephalic territories, which involves Emx3 in the dogfish, but another orthology class, Emx2, in tetrapods. In addition, the dogfish Emx2 gene shows a highly specific expression domain in the cephalic paraxial mesoderm from the end of gastrulation and throughout neurulation, which suggests a role in the segmentation of the cephalic mesoderm.

© 2002 Elsevier Science (USA)

Key Words: Emx genes; multigene family; chondrichthyans; dogfish; brain regionalization; cephalic mesoderm.

INTRODUCTION

In the last two decades, comparative molecular analyses have provided an increasing body of evidence suggesting that conserved genetic mechanisms may control the early regionalization and patterning of the central nervous system. For instance, an early division into an anterior territory, expressing Otx/otd genes, and a posterior territory, expressing Hox genes, appears highly conserved in triploblastic metazoans, including annelids, insects, and a wide range of deuterostomes (Arendt and Nübler-Jung, 1999). The conservation of additional subdivisions in the brain region has also been proposed, mainly on the basis of comparisons between Drosophila and the mouse (Arendt and Nübler-Jung, 1996; Hirth and Reichert, 1999). However, the interpretation of such large-scale comparisons can be controversial, as shown by the example of empty-spiracles-related genes. In Drosophila, the homeobox empty-spiracles (ems) gene plays a major role in the development of the brain, and its inactivation results in a gap gene phenotype, characterized by the specific deletion of the two posterior-most neuromeres, the deutocerebrum and the tritocerebrum (Hirth et al., 1995). In the mouse, two ems-related genes (called Emx1 and Emx2) have been characterized. Like their Drosophila ortholog, both are transcribed in specific neuromeres of the embryonic brain, showing a prominent expression in the dorsal telencephalon, with a sharp posterior limit at the telencephalon-diencephalon boundary. This observation has led to the suggestion that ems-related genes may be part of an ancestral genetic network controlling early forebrain regionalization (Simeone et al., 1992a,b). However, functional analyses in mice and comparative analyses in chordates have failed to bring a strong support to this conclusion. First, inactivations of Emx1 and Emx2 in mice result in relatively weak brain phenotypes, which substantially differ from the broad deletions observed in Drosophila. Only minor defects in the
corpus callosum have been observed in Emx1−/− mice, in some, but not all, genetic backgrounds (Qiu et al., 1996). Similarly, Emx2−/− mice die at birth due to defects in the urogenital system, another major site of Emx2 expression, and the brain abnormalities which they display are mainly restricted to medial pallium structures, with a reduction of the hippocampus and the absence of dentate gyrus (Miyamoto et al., 1997; Pellegrini et al., 1997; Yoshida et al., 1997). Second, an ems-related gene is present in the two protochordates studied thus far, the ascidian H. roretzi and the cephalochordate Branchiostoma floridae (Williams and Holland, 2000; Oda and Saiga, 2001). However, expression studies in the former have failed to detect any expression in the central nervous system, transcripts being only present in the epidermis.

A comparative analysis of the expression patterns displayed by Emx genes at a shorter evolutionary scale, among vertebrates, can be a first step to understand their functional evolution. Vertebrate ems-related genes (termed Emx) have thus far characterized in a wide range of osteichthyan, including the mouse, the zebrafish, Xenopus, and the chick (Morita et al., 1997; Williams and Holland, 2000). They all share a prominent dorsal telencephalic expression domain, but display various additional expression sites in restricted areas of the diencephalon and mesencephalon, or in the developing sense organs. An Emx gene, LjEmx, has also been recently characterized in the lamprey Lampetra japonica (Myojin et al., 2001). It shares a highly specific expression territory in the dorsal telencephalon with its osteichthyan counterparts (Murakami et al., 2001), but other expression features substantially differ between the two taxa. For instance, no diencephalic expression territories have been thus far described in the lamprey. LjEmx also displays a prominent expression in the mesoderm starting from neurulation stages, an expression domain which has never been reported thus far in osteichthyan (Myojin et al., 2001). In order to gain new insights into the structural and functional evolution of vertebrate Emx genes, we have characterized these genes in a shark, the dogfish Scyliorhinus canicula, which joins the chondrichthyans, a group of early emergence among gnathostomes. These data clearly indicate the presence of three Emx orthology classes among gnathostomes and provide new insights into the evolution of their expression patterns among gnathostomes.

**MATERIALS AND METHODS**

**Embryos**

Dogfish eggs were collected from freshly killed S. canicula females and kept in oxygenated sea water at 14°C until the desired stages were reached. After carefully opening the egg shell, the embryos were dissected and staged according to Ballard et al. (1993).

**cDNA Amplification and Sequence Analysis**

The strategy used to characterize Emx homologs in the dogfish S. canicula consisted in the successive amplification and sequencing of five partially overlapping DNA segments numbered from 1 to 5, following their order of amplification (Fig. 1). Fragments 1, 2, and 3 were amplified from dogfish embryonic cDNA (stage 21) by using a degenerate RT-PCR approach. Fragment 1 was first amplified by a seminested PCR, using as primer, 5'-TNYTTYACNNHTGRTC-3' and as 3' primers, 5'-YTGRAANCKRTGNCNCA-3' and 5'-ATNARCCANGRTARAA-3'. These three primers respectively correspond to the CFTIES, FYPWVL, and FYPWLI motifs conserved among osteichthyan Emx genes (cycling conditions: 95°C, 1 min; 45°C, 1 min; 72°C, 1 min; 40 cycles). Fragments 2 and 3 were then successively amplified by using nested 5' primers located in PCR fragment 1 and specific for each Emx gene identified, and degenerate 3' primers corresponding to the QKLEE and (E/D)E(D/D)DVT protein motifs (respectively, S-TCTYTCTC-NARYTYTG-3' and S-GTNACRTCDATNTCTC-3', cycling conditions: 95°C, 1 min; 45-50°C, 1 min; 72°C, 1 min; 40 cycles). Genomic DNA fragments encoding the N-terminal and C-terminal parts of each Emx protein identified (fragments 4 and 5 in Fig. 1) were obtained by inverse PCR, starting from digested and religated adult liver genomic DNA, and using specific primers chosen in the neighboring sequenced fragments (fragments 1 and 3). As a control, a cDNA fragment spanning the full-length coding region was amplified for each gene identified, in a single RT-PCR step, using specific primers located upstream of the initiation codon and
downstream from the termination codon (fragment 6 in Fig. 1). Amplified fragments were subcloned and sequenced by using standard protocols. In each case, nucleotide sequences were deduced from the consensus sequence of at least three independent clones. The three dogfish Emx sequences identified were submitted to GenBank under the following Accession Nos.: AF306635, AF306636, and AF306637.

Phylogenetic Analysis and Database Search

The following Emx protein sequences were retrieved from the GenBank database or from the following references: Mus musculus Emx1 and Emx2, Homo sapiens Emx1 and Emx2 (Pataranello et al., 1997), Gallus gallus Emx1 and Emx2 (Bell et al., 2001), Xenopus laevis Emx1 and Emx2 (Pannese et al., 1998), Danio rerio Emx1 (D32214) and Emx2 (D32215), Oryzias latipes Emx1 (A1250402) and Emx2 (A132403), L. japonica Emx (AB040759), and B. floridana EmxA (AF261146). Similarities to Emx coding sequences were found in scaffolds retrieved from http://fugu.hgmp.mrc.ac.uk/fugu/bin/clonesearch, with the following Accession Nos.: scaffold 6997 (FT:T006997), scaffold 12441 (FT:T012228), scaffold 12231 (FT:T012228), scaffold 2099 (FT:T002114), and scaffold 2099 (FT:T002099). Human Emx1 and Emx2 sequences, which are almost identical to their mouse orthologs and the chick partial Emx1 sequence, were excluded from the phylogenetic analysis. The sequences were aligned by using the ED program of the MUST package (Philippe, 1993). Similarities to Emx protein sequences were searched for in the genomic and EST sequences available from human, mouse, pufferfish, and zebrafish (NCBI human and mouse genomic BLAST pages, NCBI EST database, and Fugu Genome Consortium BLAST page) by using the tblastn algorithm (Altschul et al., 1997) by using the tblastn algorithm (Altschul et al., 1997) with X. laevis and D. rerio Emx1 and Emx2 protein sequences as inputs. Sequences displaying strong similarities were retrieved and further analyzed. Phylogenetic analyses were performed by using neighbor-joining (NJ), maximum-parsimony (MP), and maximum likelihood analyses (ML). For the NJ analyses, distance matrices were calculated according to Rzhetsky and Nei (1994) with parameter alpha set to 0.5 according to the ML estimate, and trees were constructed by using the NJ program of the MUST package. MP analyses were performed by using the branch and bound algorithm available in PAUP version 3.1.1 (Swofford, 1993), and ML was conducted by using PROML version 2.3 (Adachi and Hasegawa, 1996); the best ML tree was chosen among 2000 trees constructed by using the quick-add OTUs method with the JTT-f model of amino acid substitution. Confidence in each node was assessed by 1000 bootstrap replicates in NJ and MP analyses and by the RELL method (Kishino et al., 1990) on the 2000-ml trees constructed.

Whole-Mount Hybridization of Dogfish Embryos

Digoxigenin 11-UTP-labeled antisense RNA probes were transcribed in the presence of T7 RNA polymerase from linearized pT219R recombinants containing the whole coding region of each gene identified (fragment 6). Whole-mount hybridizations of dogfish embryos were performed according to standard protocols (Xu and Wilkinson, 1993). No cross-hybridization between the paralogous genes was observed under these conditions. After hybridization, embryos were transferred to 75% glycerol/PBT (PBT: phosphate-buffered saline, 0.1% Tween 20), and photographed or processed for sections.

Histological Sections

Following whole-mount hybridization, embryos were postfixed in 4% paraformaldehyde/PBT (4°C, 12 h), rinsed in PBS, equilibrated in 15% sucrose (4°C, overnight), embedded in 15% sucrose, 20% gelatin (37°C, overnight), frozen in liquid nitrogen, and mounted in O.C.T. embedding compound (Miles Elkhart) for cryostat sections. Then, 20-μm sections were cut and mounted onto gelatinized slides. The slides were counterstained by using Nuclear Fast Red (Vector Laboratories, Burlingame, CA), dehydrated, mounted in Eukitt (O. Kindler GmbH, Freiburg, Germany), and photographed by using Nomarski optics.

RESULTS

Identification of Three Emx Genes in the Dogfish S. canicula

To identify empty-spiracles-related genes in the dogfish S. canicula, we used a degenerate PCR-based strategy, taking advantage of the high phylogenetic conservation of protein motifs present in all osteichthyan Emx proteins (Fig. 1; see Materials and Methods). This strategy led to the characterization of three distinct cDNA fragments, containing open reading frames encoding proteins of 232, 234, and 246 amino acids in length, respectively. The three genes thus identified can be unambiguously assigned to the ems class, as shown by their comparison with Drosophila ems. First, their deduced amino acid sequences contain a homeodomain which displays a high similarity to the one of ems (12–13 differences between ems and each of the three dogfish sequences). This similarity extends to two residues located immediately upstream from the homeobox (RK) and three residues immediately downstream (KEE or EEE in the...
FIG. 3. Emx-related coding sequences in F. rubripes. (A) Scheme showing the position in the coding region of the similarities found in F. rubripes. The exon-intron organization of vertebrate Emx-coding sequences is shown in the upper line; a black box indicates the homeodomain, and arrowheads indicate intron-exon junctions. Horizontal bars delimit the region of similarity found in each sequence scaffold relative to the coding sequence. Emx1, Emx2, and Emx3-related sequences are shown in red, yellow, and blue, respectively. Numbers between brackets indicate the position of the similarity in the nucleotide sequence of each scaffold. (B) Alignment of the deduced
In order to investigate orthology relationships between the three dogfish Emx genes and the Emx1 and Emx2 genes identified in oostichthyans, we performed a first phylogenetic analysis using neighbor-joining (NJ), maximum-parsimony (MP), and maximum likelihood (ML) algorithms (Fig. 2B). The whole protein sequence was taken into account in this analysis, except for two segments, which correspond to broad insertions either in the lamprey and Amphioxus, or mammalian an sequences (positions 22–56 and 94–102 in Fig. 2A). All three reconstruction methods led to identical tree topologies (Fig. 2B). In all cases, three distinct classes can be identified, each containing a single dogfish and at least one oostichthy an Emx sequence. As expected, the four Emx2 sequences characterized in oostichthyans (MmEmx2, GgEmx2, XIEmx2, and DrEmx2) cluster within a single class (class 2), together with one of the dogfish sequences, theretof er termed ScEmx2. By contrast, a partition of oostichthy an Emx1 sequences into two distinct classes is observed. One of them (class 1) contains the mouse and Xenopus Emx1 sequences together with a second dogfish Emx1 form (which was termed ScEmx1). The zebrafish Emx1 sequence does not emerge within this class but appears closely related to the third dogfish sequence (termed ScEmx3), thus defining a third class (class 3). The partition of the gnathostome Emx genes into these three classes is supported by high bootstrap values whatever the reconstruction method used (Fig. 2B). In addition, inside each class, the branching orders of the paralogous genes are identical in the tree reconstruction methods used, and consistent with the gnathostome phylogeny, ScEmx1 and ScEmx2 displaying basal positions among gnathostome Emx1- and Emx2-related genes.

In order to study the relationships of the lamprey LjEmx and Amphioxus EmxA genes with their oostichthy an counterparts, we performed a second analysis including these sequences (Fig. 2C). The alignment segments used in this analysis were restricted to positions 2–21 and 162–312 (Fig. 2A), due to the extensive divergence of the lamprey and Amphioxus sequences with their oostichthy an counterparts outside these regions. Whatever the reconstruction method used, the monophyly of each of the three gnathostome Emx classes is recovered with good statistical supports in this analysis (Fig. 2C). In contrast, the order of emergence of the three classes and the relative position of the lamprey gene were found to vary depending on the reconstruction method used (data not shown) and could not be confidently resolved.

### Database Search of Novel Oostichthy an Emx1-, Emx2-, and Emx3-Related Sequences

In order to investigate the presence of representatives of the Emx3 class in sarcopterygians and of the Emx1 class in actinopterygians, we performed systematic searches for Emx-related sequences in the sequence databases currently available for human, mouse, and zebrafish (http://www.ncbi.nlm.nih.gov/BLAST/), and for the pufferfish Fugu rubripes (http://fugu.hgmp.mrc.ac.uk/blast/). While both Emx1 and Emx2 sequences could be readily retrieved from mouse ESTs and the available human genome data, no indication of a third, Emx3-related gene could be obtained in these species. Similarly, no other sequence than those encoding ZfEmx1 and ZfEmx2 could be found in the zebrafish sequence data. In contrast, a strong evidence for the presence of three Emx genes was obtained from the pufferfish genomic sequence. In this species, five distinct sequence scaffolds (scaffolds 6997, 2114, 2099, 12441, and 12231) containing a total of eight segments of similarity to Emx coding sequences were retrieved from the raw sequence database (Fig. 3A). In most cases, the similarities end accurately at previously reported intron–exon junctions (Figs. 3A and 3B), confirming the conservation of intron positions in chordates (Williams and Holland, 2000). The only exception concerns nucleotides [239–346] in scaffold 2099, whose deduced amino acid sequence shows a very strong similarity to part of the gnathostome Emx2 sequence encoded by the first exon (positions 170–232 in Fig. 2A). However, this similarity is abruptly interrupted upstream of position 105. We could not detect any intron
acceptor site or candidate initiation methionine in the vicinity of this truncation, and its biological significance remains unclear. Similarities to exon 2 could not be detected in this scaffold, but may be present in undetermined sequences lying between positions 964 and 1387.

To study the relationships of the Emx sequences deduced from the F. rubripes genome with their gnathostome counterparts, we included their deduced amino acid sequences in the alignment shown in Fig. 2A, together with two medaka partial sequences retrieved from GenBank as O1Emx1 and O1Emx2 (Fig. 3B; and data not shown). A third phylogenetic analysis including these sequences was performed by using ML, NJ, and MP algorithms (Fig. 3C). This analysis was restricted to the protein sequences encoded by the first exon, which vary substantially among the three Emx classes. The Fugu Emx sequences contained in scaffold 2099, which display a truncation in this region, were excluded from this comparison. The resulting phylogenetic tree confirms the partition of the gnathostome Emx sequences into three classes, each containing one of the dogfish sequences and supported by high bootstrap values (Fig. 3C). Whatever the reconstruction method used, the Fugu amino acid sequences deduced from scaffolds 6997 and 12441 emerge within the Emx1 and Emx3 classes, respectively. As to the partial Emx sequences identified in scaffold 2099, they are almost identical to the zebrafish and dogfish Emx2 sequences, which supports their assignment to the Emx2 class (Fig. 3B).

The partition of the gnathostome Emx sequences into three distinct classes is further substantiated by the sharing of residues, which appear selectively conserved within each class (shaded in Figs. 2A and 3B). Such distinctive features are observed in relatively variable parts of the molecule, but also in the homeodomain (with three Emx2-specific residues: positions 223, 241, 261; one Emx1-specific residue: position 246; and one Emx3-specific residue: position 244), or in the close vicinity of motifs conserved in Drosophila (with five Emx2-specific residues: positions 162, 168, 174, 176, and 178; and two Emx1-specific residues: positions 180 and 181).

Expression of ScEmx2 and ScEmx3 during Gastrulation and Neurulation (Stages 13-17)

**ScEmx1.** No expression of ScEmx1 was detectable in S. canicula embryos until stage 19.

**ScEmx2.** ScEmx2 transcripts were first observed in gastrulating embryos at stage 13. At this stage, ScEmx2 displays two longitudinal expression domains, symmetrically located on each side of the embryonic axis in its anterior half (Fig. 4A). This expression domain intensifies and its boundaries become more sharply defined at stage 14, when the cephalic enlargement becomes visible (Fig. 4B): at this stage, the hybridization signal is restricted to the cephalic enlargement at a prechordal level of the anterior-posterior axis. Transverse sections show that this ScEmx2 expression is confined to the lateral mesoderm, being clearly excluded from the axial mesoderm which forms a continuous layer with the adjacent ectoderm and endoderm layers (Figs. 4E and 4H). At stage 17 (neural tube closure), this mesodermal expression is present in the epithelial layer lining the first head coelomic cavity (mandibular cavity), which becomes visible on each side of the embryonic head, with a higher signal intensity in its lateral parts (Figs. 5A and 5C). A hybridization signal in the cephalic mesoderm, located rostrally to the first pharyngeal pouch, is still visible at stage 19 but faints afterwards (Fig. 6A). The signal which is observed at later stages in the head mesenchyme, on each side of the telencephalon and diencephalon (Figs. 7F and 7G) clearly corresponds to a distinct phase of expression. No expression could be detected in the ectoderm until stage 19.

**ScEmx3.** ScEmx3 expression was first detected at stage 14. At this stage, a strong hybridization signal appears as two symmetrical lateral stripes lining the anterior border of the head enlargement (Fig. 4C). As shown by transverse sections (Fig. 4F), this signal is restricted to the neural plate, with a sharp lateral boundary at the border between the presumptive neural ectoderm and skin ectoderm. Transcripts are also present in the presumptive skin ectoderm, but at more posterior levels (Figs. 4C and 4I). At later stages, the anterior domains of expression fuse dorsally upon neural tube closure, resulting in a broad dorsal expression domain in the anteriormost part of the neural tube (stage 17; Figs. 5B and 5D). A faint expression in the first pharyngeal pouch is also first observed at stage 17 (Fig. 5B).

Expression of ScEmx1, ScEmx2, and ScEmx3 in the Embryonic Brain (Stages 19-22; 2.5-cm Embryos)

**ScEmx1.** ScEmx1 transcription was first detectable at low levels in the developing brain at stage 19 (data not shown). During stages 21-22, a faint hybridization signal was observed dorsally, in the caudal part of the telencephalon (Fig. 7A). As shown by sagittal sections, its posterior border of expression is located immediately anterior to the velum transverse, which corresponds dorsally to the dorsal telencephalon-diencephalan boundary (Fig. 7M). Expression declines dorsally, being almost undetectable along the dorsal midline. In 2.5-cm embryos, transcripts display a widely spread distribution, spanning the dorsal telencephalon, the infundibulum, and the dorsal mesencephalon. However, they remain excluded from the dorsal diencephalon (Fig. 8A).

**ScEmx2.** ScEmx2 transcription in the embryonic brain was first detected at stage 19, in the ventral half of the prosencephalon, posterior to the optic evagination (Figs. 6A-6C). This ventral expression domain persists at stage 22, with transcripts present in the ventral part of the diencephalon, posterior to the optic stalk, but excluded from the floor plate and adjacent cells (Fig. 7H). At this stage, another domain of expression is observed in the caudal part of the dorsal telencephalon. This expression domain is very similar to the one displayed by ScEmx1,
FIG. 4. Expression of ScEmx2 and ScEmx3 during gastrulation and the beginning of neurulation. (A–C) Dorsal views of stage 13 (A) or stage 14 (B, C) S. canicula embryos, after whole-mount hybridization using ScEmx2 (A, B) or ScEmx3 (C) antisense RNA probes. Anterior is to the top. (D–I) Transverse half-sections of the embryos shown in (A–C) respectively. Medial is to the left. Thin lines in (A–C) indicate the level of the sections. ect, ectoderm; m, mesoderm; end, endoderm; np, neural plate; ep, presumptive epidermis; nt, notochord.

FIG. 5. Expression of ScEmx2 and ScEmx3 stage 17 S. canicula embryos. (A, B) Lateral views (anterior to the top) of stage 17 embryos after ISH using ScEmx2 (A) and ScEmx3 (B) probes. (C, D) Frontal (C) and sagittal (D) sections of the embryos shown in (A) and (B), respectively (anterior to the left). A thin line in (A) indicates the plane of the section shown in (C). A black arrowhead indicates ScEmx3 expression in the first branchial pouch. mdc, mandibular cavity.
albeit a significantly higher signal intensity was reproducibly observed (Figs. 7E and 7G). At later stages, in 2.5-cm embryos, prominent signals are present in the posterior infundibulum and the dorsal telencephalon with a sharp posterior border at the telencephalon–diencephalon boundary (Fig. 8B). In the telencephalon, the transcripts are excluded from the dorsal midline. A faint and diffuse labeling is also present in the ventral mesencephalon. No transcripts could be detected in the dorsal diencephalon.

ScEmx3. The boundaries of ScEmx3 expression domain in the anterior part of the neural tube become sharply defined at stage 19 (Figs. 6D–6G). At this stage, ScEmx3 displays a broad expression domain, extending over the whole dorsal prosencephalon, with a sharp posterior border immediately anterior to the prosencephalon–mesencephalon boundary. Dorsally, a transverse line of expression at the mesencephalon–metencephalon limit can also be observed (Fig. 6D). Both of these expression domains persist during stages 21 and 22 (Fig. 7I). At these stages, ScEmx3 expression domain in the forebrain is mainly restricted to the dorsal telencephalon but clearly extends dorsally beyond the velum transverse, into diencephalic territories (Figs. 7J–7L and 7O). However, in the absence of morphological landmarks, its boundaries remain difficult to define more precisely. In 2.5-cm embryos, ScEmx3 shows a prominent expression spanning the telencephalic evaginations and the dorsal midline. At this stage, the signal is completely excluded from more posterior, diencephalic areas (Fig. 8C).

Other Sites of Expression

ScEmx genes are also transcribed in other parts of stage 22 embryos, including the otic and olfactory placodes, the pharyngeal arches, and the primordium of the urogenital system. All three ScEmx genes are expressed in the olfactory placodes, albeit with nonsuperimposable expression domains. While ScEmx1 and ScEmx2 display relatively faint signals, confined to a restricted area located to the middle of the epithelium thickening (Figs. 7A, 7B, 7E, and 7F), ScEmx3 domain of expression in the olfactory placodes is much broader, extending dorsally into the epidermis, up to the midline (Figs. 7I and 7J). In the otocyst, only ScEmx2 is transcribed, with a signal restricted to the mediodorsal quadrant. All four anterior pharyngeal arches also express different combinations of ScEmx genes at stage 22. ScEmx1 transcripts remain undetectable until stage 22. At this stage, they are restricted to a posterior ectodermal region, which is located halfway along the proximodistal axis (Fig. 7A; sections not shown). ScEmx2 and ScEmx3 expressions in the pharyngeal arches appear earlier, from stage 19 on. At stage 22, ScEmx2 expression domain only concerns a restricted mesodermal area of the mandibular and hyoid arches, while ScEmx3 expression is detected in the whole ectoderm of all four visceral arches (Figs. 7E and 7I; sections not shown). Finally, all three genes are expressed in the urogenital system of stage 22 embryos, with a very low signal intensity for ScEmx3.

DISCUSSION

Three Orthology Classes among Gnathostome Emx Genes

The monophyly of osteichthyan Emx2 genes has been undisputed thus far, but less clear results have been obtained concerning Emx1 genes in previous phylogenetic
analyses (Patarnello et al., 1997; Williams and Holland, 2000). While the tetrapod Emx1 genes were always found clustered in a single clade with high support values, the relative position of the zebrafish Emx1 gene could not be confidently resolved. These ambiguous results led to the suggestion that either this gene had undergone a particularly high rate of evolution in actinopterygians, or it was a member of another class, thus far unidentified in sarcopterygians. The comparison of the three Emx sequences characterized in S. canicula with their osteichthyan counterparts provides a strong support in favor of the latter hypothesis. Whatever the method and the parameters used in the phylogenetic analysis, the gnathostome Emx genes appear partitioned into three monophyletic groups, each containing one of the three dogfish sequences, and either the sarcopterygian Emx1 (Emx1 class), the osteichthyan Emx2 (Emx2 class), or the zebrafish (and medaka) Emx1 sequences (Emx3 class). We conclude from these results that three Emx genes were present in the lineage of gnathostomes prior to their radiation and that the three groups identified in our analysis derive from these ancestral genes. The Emx1 genes isolated in the zebrafish (or medaka) and in the tetrapods therefore belong to distinct orthology classes.

As shown by the pufferfish genome analysis, all three orthology classes were clearly present in the last common ancestor of teleosts and it will be of great interest to investigate the presence of an Emx1-related gene in the zebrafish. In sarcopterygians, the presence of an Emx3-related gene remains an unanswered question. The presence of an additional Emx-related gene, which could compensate for some of the functions of its paralogs, could account for the relatively weak Emx1/− and Emx2/− phenotypes observed in mice. However, while both Emx1 and Emx2 coding sequences could be readily retrieved from mouse ESTs and the available human genome data, no indication of a third, Emx3-related gene could be obtained in these species. We therefore favor the hypothesis that Emx3-related genes have been lost during the evolution of sarcopterygians, prior to the mammalian radiation.

Both Telencephalic and Diencephalic Emx Expression Sites Are Conserved among All Gnathostomes

Dorsal telencephalon. During early organogenesis, all osteichthyan Emx genes share broad dorsal expression domains in the telencephalon, showing sharp posterior borders at the telencephalon–diencephalon boundary (Simeone et al., 1992a,b; Pannese et al., 1998; Bell et al., 2001; Morita et al., 1995). Similarly, when the major subdivisions of the lamprey brain become visible (stage 26), the L]Emx gene identified in L. japonica appears restricted to a dorsal subdomain of the telencephalon. These data have suggested that Emx genes could be part of a vertebrate ancestral genetic network, which controls the early regionalization of the telencephalon and is highly conserved in this taxon (Murakami et al., 2001). Our analysis of Emx genes in a chondrichthyan supports this conclusion. Like their vertebrate counterparts, the three dogfish Emx genes are transcribed in broad domains of the dorsal telencephalon with a sharp posterior border at the telencephalon–diencephalon boundary when the cerebral vesicles become visible. The high conservation of this Emx expression border in vertebrates is consistent with an involvement of Emx genes in its establishment or maintenance. However, no such role has been clearly demonstrated thus far in Emx1 or Emx2 single mutant mice, possibly due to functional redundancies between the two mouse genes (Yoshida et al., 1997; Qiu et al., 1996; Pellegrini et al., 1996). In line with this possibility, the Gli3 mutation, which abolishes the expression of both Emx1 and Emx2 in 10.0-dpc mouse embryos, results in morphological and molecular abnormalities at the demarcation between the dorsal diencephalon and the telencephalon (Theil et al., 1999). Analyses of Emx1 and Emx2 double-mutant mice will be important to evaluate possible roles of Emx genes in the morphogenesis of this boundary.

Dorsal diencephalon. While the ScEmx3 expression domain is restricted to the dorsal telencephalon in 2.5-cm embryos, it initially spans more caudal territories. In an initial phase, ScEmx3 expression pattern is dynamic, spanning a broad dorsal territory with posterior borders coincident with the prosencephalon/mesencephalon boundary when it becomes visible, and later spanning the dorsal telencephalon and anterior-most parts of the dorsal diencephalon. This early phase of expression, which is first detected at the onset of neurulation, is not observed for the paralogous genes ScEmx1 and ScEmx2 and has not been described in the lamprey thus far. In contrast, in the mouse as well as in the chick, Emx2 (but not Emx1) is first transcribed at early neurula stages, in a broad dorsal territory initially showing a caudal border at the prosencephalon/mesencephalon limit and later regressing inside the diencephalon (Simeone et al., 1992; Bell et al., 2001). Similarly, in its earliest phase of expression, from early to midneurula stages, the Xenopus XlEmx2 gene is transcribed in a broad dorsal domain of the forebrain, comprising both telencephalic and diencephalic presumptive territories, while XlEmx1 transcription appears restricted to presumptive telencephalic territories (Pannese et al., 1998). Identical conclusions have been suggested in the zebrafish, although the boundaries of ZfEmx1 and ZfEmx2 expression domains were not precisely mapped in this species (Morita et al., 1995). These data suggest that an early Emx activity in the dorsal prosencephalon, including presumptive diencephalic territories, may be highly conserved among gnathostomes, suggesting an essential role in the specification of these territories. Despite this conservation, analyses of Emx1/− or Emx2/− mutant mice have failed to reveal such a role. However, studies of Emx2/−/Otx2/− embryos have recently shown that Emx2 and Otx2 act cooperatively to specify dorsal diencephalic territories at early stages of the neural plate development (Suda et al., 2001). Such a cooperative interaction between Emx and Otx2 genes may be conserved in chondrichthyans, since in the dogfish as in osteichthyans, Otx2 is...
transcribed in the prospecencephalon and mesencephalon starting from the end of gastrulation (Mazan et al., 2000 and data not shown). In this hypothesis, this interaction which involves Emx2 in sarcopterygians, clearly involves Emx3, a member of another orthology class, in the dogfish.

Ventral diencephalon. In the dogfish, ScEmx2, but no other paralog, shows a strong expression in the ventral diencephalon at the end of neurulation, which becomes morphologically visible. A very similar Emx2 expression has been described in the mouse, with a prominent expression in posterior parts of the ventral diencephalon starting from 9.5 dpc (Simeone et al., 1992a,b) and in the posterior hypophysis at later stages (13.5 dpc). Emx2 expressions in the ventral diencephalon have also been reported in the chick (stage HH12–HH13; Bel et al., 2001), Xenopus (stage 37; Pannese et al., 1998) and zebrafish (24 h of development; Morita et al., 1995). These data suggest that an Emx2 expression in the ventral diencephalon may be highly conserved in gnathostomes. However, the biological significance of this conservation remains unclear, since the derivatives of the ventral diencephalon display no obvious defect either in Emx2+/− or Emx2−/−/Otx2−/− embryos.

Conclusion. Taken together, our data indicate that not only telencephalic, but also diencephalic Emx expression sites are highly conserved among gnathostomes at early stages of brain development. This conservation may seem paradoxical since relatively subtle defects have been thus far observed in mice carrying either Emx1−/− or Emx2−/− mutations. However, as suggested by analyses of Gli3 or Emx2−/−/Otx2−/− mutations, it could reflect a selective pressure for the maintenance of functional redundancies between the paralogous genes and possibly other homeodomain genes. An alternative, although not exclusive interpretation may be that some abnormalities may escape detection in laboratory conditions and provide a selective advantage at the population level (Cooke et al., 1997). This may be particularly true for Emx genes, which in mammals, are involved in the arealization of the cortex (Bishop et al., 2000; Mallamaci et al., 2000) and in the morphogenesis of the hippocampus, a brain structure known to play an important role in memory processes (Pellegrini et al., 1996; Yoshida et al., 1997; Tole et al., 2000).

ScEmx2 Expression in the Mesoderm Suggests a Role in the Segmentation of the Cephalic Mesoderm

At the trunk level, the segmentation of the vertebrate paraxial mesoderm into somites has been extensively documented, and the underlying genetic and cellular mechanisms largely elucidated (reviewed in Pourquié, 2001). In contrast, the vertebrate cranial paraxial mesoderm, located anterior to the segmental plate, never condenses into somites and its organization has remained a matter of controversy for more than a century. Classical interpretations have proposed that it may consist of several segmental units, corresponding to vestigial forms of genuine segments present in a vertebrate ancestor (reviewed in Kuratani et al., 1999; Holland, 2000). These segmental theories were largely influenced by morphological analyses of amphioxus embryos, which display a fully segmented organization of the paraxial mesoderm along the whole body length, including its most rostral aspects, but also of shark embryos, which exhibit three to four pairs of tandemly arranged cavities in the cephalic paraxial mesoderm. Different kinds of experimental evidence supporting these views have been more recently obtained. First, a strong support for a homology between the segmented anterior-most paraxial mesoderm present in amphioxus embryos, and the cephalic paraxial mesoderm lying rostral to rhombomere 5 in vertebrates, has emerged from comparative analyses of Hox expression patterns in these species (Holland, 2000). Second, a transient segmental pattern has been described in the cephalic paraxial mesoderm of a wide range of vertebrates, on the basis of scanning electron microscopy observations. Four tandemly arranged units, corresponding, from rostral to caudal levels, to the premandibular mesoderm, the mandibular mesoderm, the hyoid mesoderm, and the so-called somite 0 (a somite-like unit showing an incomplete segmental cleft at its anterior border), have thus been described at neural plate stages in lampreys, sharks, and amphibians (Kuratani et al., 1999; Kuratani and Horigome, 2000; Jacobson and Meier, 1984). An early subdivision of the cephalic paraxial mesoderm in seven somitomeres has also been reported in anniotes as well as in teleosts, suggesting that modifications of an ancestral segmentation pattern may have independently occurred in these two groups (reviewed in Jacobson, 1988). Finally, a regionalization of the cephalic paraxial mesoderm has been unambiguously shown by cell lineage analyses in mouse and in chick embryos (Couly et al., 1992; Trainer et al., 1994). In the head region, the fate of paraxial mesoderm cells is predominantly myogenic but clearly differs along the rostrocaudal axis, cells located at prosencephalon and mesencephalon levels contributing to different oculomotor muscles, while those located at the rhombencephalon level colonize jaw muscles. However, in the absence of morphological landmarks, it is difficult to formally assign these different structures to individual segments. Furthermore, no genetic evidence supporting an early segmental pattern of the cephalic paraxial mesoderm has been obtained thus far, and the genetic mechanisms, which control its regionalization, remain completely unknown.

In this context, the highly spatially restricted mesoderm expression of ScEmx2 strongly suggests its involvement in the early specification of a well-individualized component of the cephalic paraxial mesoderm, thus providing a strong genetic argument in favor of its segmental organization in the dogfish. This component most likely corresponds to the presumptive mandibular mesoderm, since the labeling appears restricted to cells that delimit the mandibular cavity when it becomes visible. Whether this Emx expression corresponds to a derived feature of sharks or a primitive
character of gnathostomes or even vertebrates, is an important, but unresolved issue. Thus far, no Emx expression in the paraxial mesoderm has been described either in osteichthyans, or in protochordates. In contrast, the LjEmx gene isolated in *L. japonica* shares an early mesodermal expression with ScEmx2, suggesting that this expression feature...
Conclusion: Functional Evolution of Emx Genes in Vertebrates

It is currently widely accepted that massive gene duplications have occurred early in the vertebrate lineage, and previous analyses have led to the conclusion that Emx genes provide a novel example of these genetic events (Williams and Holland, 2000). Our analysis supports this conclusion since we show here that the duplication of Emx genes has predated the splitting of chondrichthyans and osteichthyans, which is consistent with the chronologies inferred from other genetic systems (Holland et al., 1994).

Whether the diverse sites of expression shown by gnathostome Emx genes are primitive among chordates remains an open question. The posterior epidermis during neurulation, where no expression of osteichthyan or lamprey Emx genes has been described thus far, is actually the only expression site shared by the ascidian Emx gene, one of the dogfish gene (Emx3), and possibly Drosophila ems (Oda and Saiga, 2001). Its biological significance is unclear. However, comparisons of the expression patterns displayed by the three dogfish paralogous genes and their osteichthyan counterparts provide insights into those already present in the ancestral Emx gene, which was duplicated in the vertebrate lineage to generate the three extant Emx1, Emx2, and Emx3 orthology classes. Our data suggest that this ancestral gene already showed a complex expression pattern, including the dorsal prosencephalon during neurulation, an expression site shared by the dogfish Emx3 and the tetrapod Emx2 genes, the dorsal telencephalon during early organogenesis, where LjEmx and all gnathostome Emx genes are transcribed, but also the olfactory placodes, a site of ScEmx2 and ScEmx3 expression (only Emx2 in tetrapods), and the urogenital system, a site of ScEmx1, ScEmx2, and ScEmx3 expression (Emx2 in the mouse). These sites of expression have subsequently been retained by different paralogs in different taxa, as proposed in the duplication–degeneration–complementation model of multigene family evolution (Force et al., 1999). As a consequence, gnathostome Emx genes provide a novel example of “shuffling” of expression domains, and possibly function, between paralogues, such as those observed between the zebrafish Hoxa-1 and the mouse Hoxb-1 genes (McClelland et al., 2001), or the mouse and chick slug and snail genes (Sefton et al., 1998). This does not preclude possible functionalizations of one orthology class for other Emx functions. For instance, in the dogfish as in osteichthyans, Emx2, but no other paralog, is expressed in the otic vesicle and the ventral diencephalon during neurulation. The identification of residues, which are selectively conserved in each orthology class, also suggests that different structural constraints may act on Emx1, Emx2, and Emx3 proteins. In this respect, it is noticeable that a number of Emx2-specific residues are located in the immediate proximity of a protein motif also conserved in Drosophila ems (FYPWM). In Hox proteins, a related protein sequence (YPWM), similarly located at a short distance upstream the homeodomain, is known to be an important component of protein–protein interactions which increase the specificity and affinity of Hox–DNA binding (Passner et al., 1999). Taken together, these data support the idea that the three paralogous proteins may be involved in different protein–protein interactions, whose affinity and specificity could be modulated by the residues selectively conserved in each class. Combinations of functional and comparative approaches will be necessary to gain further insights into the molecular mechanisms underlying the diversification of Emx genes in gnathostomes.
ACKNOWLEDGMENTS

The F. rubripes sequences were provided freely by the Fugu Genome Consortium for use in this publication only. We thank N. Narama for her help in typing the manuscript, and M. Pradels for her excellent technical assistance. We are grateful to M. and Mme. Le Maitre and all the Pécheries d’Armoricque staff for their help in obtaining dogfish embryos. This work was supported by grants from the Centre National de la Recherche Scientifique, the Université Paris-Sud, the Fondation de la Recherche Médicale, and by fellowships from the Ministère de la Recherche et de la Technologie (to Y.D., T. S.-S., and J.-L.P.).

REFERENCES


Received for publication March 7, 2002
Revised April 16, 2002
Accepted April 17, 2002
Published online June 7, 2002