Evolution of *Emx* genes and brain development in vertebrates

TOMASO PATARNELLO¹, LUCA BARGELLONI¹, EDOARDO BONCINELLI², FABIO SPADA², MARIA PANNESE² AND VANIA BROCCOLI²

¹Department of Biology, University of Padova, Via G. Colombo 2, I-35131 Padova, Italy
²DIBIT—Istituto Scientifico H. S. Raffaele, Via Olgettina 60, I-20132 Milano, Italy

SUMMARY

*Emx1* and *Emx2* genes are known to be involved in mammalian forebrain development. In order to investigate the evolution of the *Emx* gene family in vertebrates, a phylogenetic analysis was carried out on the *Emx* genes sequenced in man, mice, frogs, coelacanths and zebrafish. The results demonstrated the existence of two clades (*Emx1* and *Emx2*), each grouping one of the two genes of the investigated taxa. The only exception was the zebrafish *Emx1*-like gene which turned out to be a sister group to both the *Emx1* and *Emx2* clusters. Such striking sequence divergence observed for the zebrafish *Emx1*-like gene could indicate that it is not orthologous to the other *Emx1* genes, and therefore, in vertebrates there must be three *Emx* genes. Alternatively, if the zebrafish *emx1* gene is orthologous to the tetrapod one, it must have undergone to strong diversifying selection.

1. INTRODUCTION

Homeotic genes are known to play a major role in controlling body-plan development by specifying the identity of different regions along the body axes (Kenyon 1994). In vertebrates, the majority of these genes have been identified by searching for *Drosophila* homologues; several such genes are homeo box-containing genes encoding homeoproteins which act as transcription factors. Their sequence comparison and expression pattern reveal striking evolutionary conservation. Much is known about the cluster of regulatory genes specifying the regional identity in the trunk of both invertebrates and vertebrates (Carroll 1995). The HOX clusters have been extensively investigated in many different organisms and comparative studies faced the difficulty of establishing gene orthologies among species (Valentine et al. 1996). Nevertheless, the common features of these clusters are their striking conservation along evolution from nematodes to mammals, suggesting a common origin in metazoan phyla as old as 600 Ma or even older (Wray et al. 1996).

The genetic control of head development is less well documented. Mouse and human forebrain organization appears to be partly controlled by four genes of the *Otx* and *Emx* families. These families are related to genes that, in *Drosophila melanogaster*, are expressed during the formation of the most anterior parts of the body (orthodenticle, *otd*; empty spiracles, *ems*) (Simeone et al. 1992). These genes show a spatiotemporal nested expression (*Otx2* > *Otx1* > *Emx2* > *Emx1*), and identify specific regions in the mesencephalon and telencephalon in vertebrates. In particular, the expression domain of *Emx1* and *Emx2* in humans and mice is restricted to the presumptive cerebral cortex and the olfactory bulbs (Simeone et al. 1993). The key role of the *Emx* genes in forebrain formation was confirmed by investigations which reported an association of *Emx1* and *Emx2* mutations with severe alterations of the cerebral cortex in humans (Brunelli et al. 1996) and mice (Qiu et al. 1996). Recently, homeotic genes involved in brain development were also investigated in fish. Two homeoproteins were reported to be expressed in the developing brain of zebrafish (*Danio rerio*) (Morita et al. 1995). The two genes encoding these homeoproteins were named *ems1* and *ems2* since their expression pattern and their sequences were considered very similar to those observed in human and murine *Emx*, suggesting that they might be homologous genes. To test this hypothesis and, more generally, to trace the evolution of vertebrate homeobox-containing genes having an *Emx*-like structure and function, we compared the complete amino acid sequence of the two *Emx* forms derived from cDNAs of humans, mice, frogs, and zebrafish.

2. MATERIAL AND METHODS

The frog *Emx* sequences were obtained by screening a *Xenopus laevis* cDNA library. Approximately $2 \times 10^6$ PFU (plaque forming units) of an *zgt11* cDNA library, prepared from stage 24–25 *Xenopus* embryos (kindly provided...
by I. Dawid), were screened at low stringency (43% formamide, 5X SSC, 0.5% SDS, 0.1 mg ml⁻¹ denatured salmon sperm DNA, 37 °C), by hybridization with a murine Emx1 and Emx2 probe. Washing was carried out in 1X SSC, 0.2% SDS at 55 °C. Several clones were purified and phage DNA was subcloned in pGEM3 (Promega), and sequenced. All the analysed clones belong to two types, that we called x-emx1 and x-emx2. The Xenopus x-emx1-deduced peptide sequence is 234 amino acid, while x-emx2 encodes a protein of 243 amino acids.

*Lateimeria* Emx1 and Emx2 were obtained by a PCR approach using the following degenerate primers:

- 3' T-emx1F: CAGGTA(G)AAA(G)GTTTGG(C)TTT(C)CA;
- 3' T-emx1R: TAATCGTG(T)T(G)GAGGTGAC(G)GTC.

The PCR products were cloned and sequenced according to the standard procedure.

Human, mouse, and zebrafish Emx amino acid sequences were derived from the nucleotide sequences deposited in Genbank under accession numbers X68879-X68882, D32214, and D32215. A total of 267 amino acids were aligned by using the computer program CLUSTAL W. The neighbour-joining (NJ) method as implemented in MEGA (Kumar et al. 1993). When the coelacanth was included in the analysis, 50 amino acids were aligned for all species; this shorter data set spanned part of the homeodomain and part of the 3' end of the gene (alignments are available upon request to T.P.).

The constancy of the substitution rate among sequences was tested within and between clades using the method described by Takezaki et al. (1995), and implemented in a computer program kindly provided by the authors. The reference tree that we used for the rate analysis was the tree reported in figure 1a.

### 3. RESULTS AND DISCUSSION

The phylogenetic reconstruction clustered all Emx2 in a monophyletic group, whereas Emx1 genes appear clearly paralogous, since the zebrafish Emx1-like gene (z-emx1) is not part of the clade which includes the Emx1 of the other species (figure 1a). Surprisingly, z-emx1 appears to be the sister group to both the Emx1 and Emx2 clades. This branching pattern, which is well supported by the bootstrap analyses, indicates a significant divergence between the Emx1 of tetrapods and that of fishes. Within the Emx2 clade the investigated sequences appear less divergent, the only unexpected position is that of *Xenopus* (x-emx2), which gave rise to the more basal branch. By means of a 'cluster test' (Takezaki et al. 1995), we tested the rate homogeneity within the Emx2 clade, and the results showed that x-emx2 has an amino acid substitution rate significantly faster than that of the other Emx2 sequences. Similarly, we tested the rate homogeneity for the Emx1 sequences; this was only possible by forcing the position of z-emx1 into the Emx1 clade (therefore assuming that z-emx1 is orthologous to the other Emx1 genes), and using the Emx2 clade as an outgroup. This relative-rate test showed that z-emx1 has a significantly faster substitution rate than the remaining Emx1 sequences.

The surprising phylogenetic position of z-emx1 prompted us to include the coelacanth (*Latimeria chlorocephalus*) in the analysis in order to investigate whether, as within bony fish, the difference observed for the Emx1 gene also holds between Actinopterygii (the ray-finned fishes, which includes the zebrafish), and Sarcopterygii (the lobe-finned fish, including the coelacanth—one of the possible sister groups to the tetrapods’ lineage). We were able to sequence two coelacanth *Emx* genes that clustered with the Emx1 clade and the Emx2, respectively (figure 1b). In particular, *Latimeria* Emx1 gave rise to the sister group of the tetrapod lineage, clearly distinct from the zebrafish Emx1-like gene (figure 1b). This tree topology indicates that the striking differentiation observed for z-emx1, as...
The orthologous gene of tetrapods is due to different selective pressures. Organisms responded to selection pressures by selective enlargement of parts of the brain that enabled a species to occupy an adaptive niche with special success (Jerison 1973). On the other hand, the coelacanth forebrain shows an ‘inverted’ structure like the forebrain of tetrapods. It appears to be ‘non-specialized’, and has been described as an ‘early vertebrate pattern’ very much resembling that of amphibians and reptiles (Millot & Anthony 1966). The anatomical analysis of the brain architecture of extant (Latimeria) and extinct members of the Crossopterygii order (Nesides tmnditi) confirms that there are no remarkably expanded forebrain structures in this fish. This suggests that the coelacanth brain does not differ much from the brain structure of the common ancestor it shares with tetrapods. Considering that Actinopterygii represent an early lineage in vertebrate evolution, the differentiation of the forebrain structure between Actinopterygii and Sarcopterygii should be as old as the vertebrate radiation that is approximately dated as being in the Devonian period, about 400 Ma.

It is open to speculation whether the Emsx gene has played some role in vertebrate radiation by controlling forebrain development, this is obviously an extremely complex process which involves many genes other than Emsx, such as Otx, Nks-2.2, Shh, and Wnt, to name but a few (Duboule 1994; Joyner 1996). In any event, if the differences in the telencephalon structure observed between ray-finned fish and tetrapods are the consequence of adaptive processes responding to niche invasion, then the surprising phylogenetic divergence among Emsx genes could be the result of evolutionary forces that have promoted the adaptation of ray-finned fish to a variety of aquatic niches, by selectively enlarging specific telencephalic structure; on the contrary, the unspecialized brain of lobe-finned fish might have been one of the key elements that allowed this lineage to experience new environments and ultimately to invade land.

We thank Dr I. Dawid for providing us with the Xenopus embryo cDNA library, we also acknowledge Dr A. Meyer for providing us with the coelacanth DNA. The present work was supported by the Italian Telethon.

REFERENCES


Kenyon, C. 1994 If birds can fly why can’t we? Homeotic genes and evolution, Cell 78, 175–180.

Kumar, S., Tamura, K. & Nei, M. 1993 MEGA—molecular evolutionary genetics analysis, version 1.01, University Park, PA.


Swofford, D. L. 1993 *Phylogenetic analysis using parsimony (PAUP)*. Illinois Natural History Survey, Champaign, IL.


Received 21 July 1997; accepted 28 August 1997