Evolution of *Emx* genes and brain development in vertebrates

TOMASO PATARNELLO[1]*, LUCA BARGELLONI[1], EDOARDO BONCINELLI[2], FABIO SPADA[2], MARIA PANNESE[2] AND VANIA BROCCOLI[2]

[1]*Department of Biology, University of Padova, Via G. Colombo 3, I-35131 Padova, Italy*
[2]*DIBIT—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 is 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 *at 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 *emx1* and *emx2* 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 homeo box-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 2×10^6 PFU(plaque forming units) of an λgt11 cDNA library, prepared from stage 24–25 *Xenopus* embryos (kindly provided

*Author for correspondence (patarnel@civ.bio.unipd.it).

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.

Latimeria 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: TAATCGTC(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 (Thompson *et al.* 1994); gaps and ambiguous alignments were excluded from the analysis. The phylogenetic reconstruction was generated on the alignment of 167 residues by 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 1*a*.

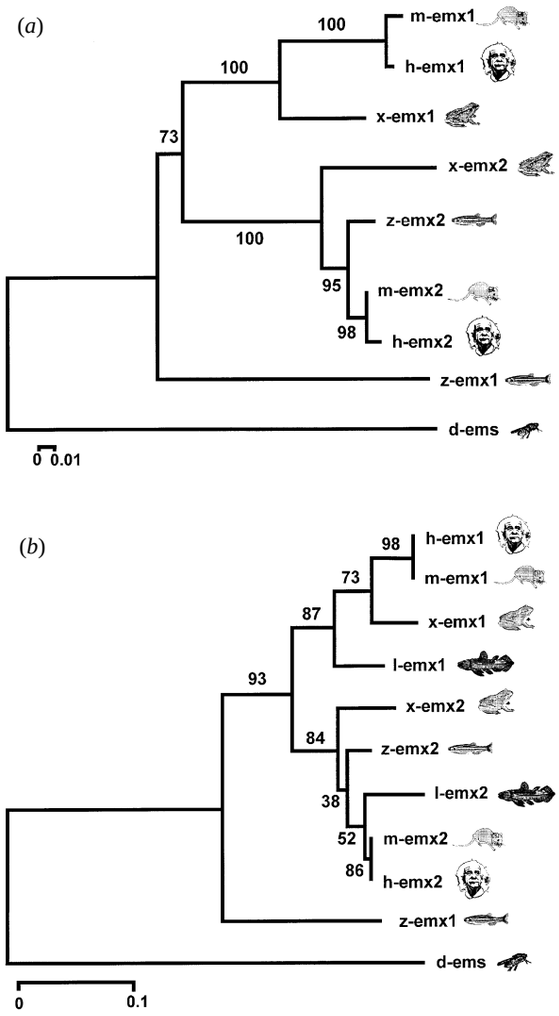

Figure 1. (*a*) Phylogenetic tree obtained by the neighbour-joining (NJ) method (Saitou & Nei 1987), as implemented in MEGA (Kumar *et al.* 1993). The tree was built on the alignment of the complete amino acid sequence derived from the nucleotide sequences of human, mouse, frog and zebrafish *Emx* genes. (*b*) NJ tree including the derived partial amino acid sequence of *L. chalumnae* (*l-emx1*, *l-emx2*). The bootstrap values (1000 replicates) are reported above each branch in both figure 1*a* and *b*. The same tree topologies were obtained by the parsimony analysis generated by PAUP (Swofford 1993). The sequence of *Drosophila melanogaster* empty spiracles (*d-ems*) was used as an outgroup in the phylogenetic reconstruction.

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 1*a*). 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 chalumnae*) 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 1*b*). In particular, *Latimeria Emx1* gave rise to the sister group of the tetrapod lineage, clearly distinct from the zebrafish *Emx1*-like gene (figure 1*b*). This tree topology indicates that the striking differentiation observed for *z-emx1,* as

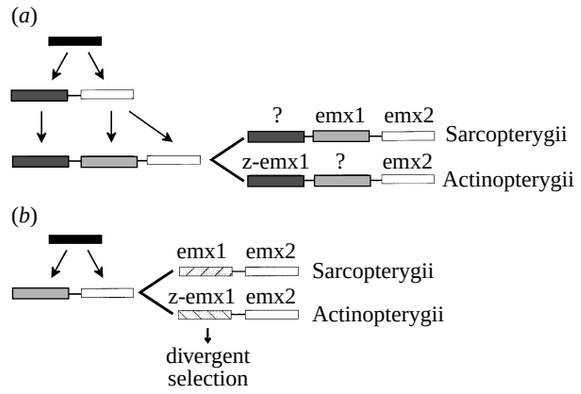

Figure 2. Alternative hypotheses for the divergent *z-emx1*: (*a*) the ancestral gene was duplicated twice before the Actinopterygii–Sarcopterygii split (giving rise to three different genes), *z-emx1* and the tetrapod *Emx1* are paralogous genes; and (*b*) there was a single duplication of the ancestral gene before the split of the Actinopterygii and the Sarcopterygii (the divergence between the zebrafish *emx1* (*z-emx1*) and the orthologous gene of tetrapods is due to different selective pressures).

compared to the other vertebrate *Emx1* genes, does not hold for the coelacanth one.

These results might have two alternative explanations, one is that the *Emx* ancestral gene was duplicated twice before the divergence of the Actinopterygii and Sarcopterygii (thus producing three distinct genes as summarized in figure 2*a*); if this was the case, then the orthologous genes (zebrafish *emx1* in tetrapods, and tetrapod *Emx* in zebrafish, respectively) were lost in the two lineages (figure 2*a*); though it cannot be excluded as it has not yet been identified.

A more parsimonious explanation in terms of gene duplication is based on the assumption that the *Emx1* of tetrapods and *z-emx1* are actually orthologous genes. In this case their remarkable divergence might be the effect of selective forces which have acted during their independent evolution. Considering the pivotal role of the *Emx* genes in forebrain development, it seems conceivable that such genetic divergence might, at least partly, account for the large differences in the forebrain structure and organization existing between ray-finned fish and the higher vertebrates. These anatomical differences were described in great detail by H. J. Jerison in an evolutionary study of the brain structure of vertebrates (Jerison 1973). In fact, the anatomy of the actinopterygian forebrain is remarkably diverse, as compared with the structure of the other vertebrates, being embryologically 'everted' rather than 'inverted'. This peculiar architecture indicates an extreme specialization of the telencephalon of actinopterygians, which accounts for their ability to occupy a great variety of niches. In this respect, it was suggested that 'the actinopterygians (which is the dominant group of fish at the present time with more than 20 000 species) 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 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 shmidti*) 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 *Emx1* 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 *Emx*, such as *Otx*, *Nkx* -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 *Emx1* 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

Brunelli, S., Faiella, A., Capra, V., Nigro, V., Simeone, A., Cama, A. & Boncinelli, E. 1996 Germline mutations in the homeobox gene *Emx2* in patients with severe schizencephaly. *Nat. Gen.* **12**, 94–96.

Carroll, S. B. 1995 Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479.

Duboule, D. (ed.) 1994 *Guidebook to the homeobox genes*. New York: Oxford University Press.

Jerison, H. J. 1973 *Evolution of the brain and intelligence*. New York: Academic Press.

Joyner, A. L. 1996 *Engrailed*, *Wnt*, and *Pax* genes regulate midbrain–hindbrain development. *Trends Genet.* **12**, 15–20.

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.

Millot, J. & Anthony, J. 1966 *Anatomie de* Latimeria chalumnae. II. Paris: CNRS.

Morita, T., Hiroyuki, N., Yuji, K., Hisashi, M. & Masayoshi, M. 1995 Differential expression of two zebrafish *emx* homeoprotein mRNAs in the developing brain. *Neurosci. Lett.* **198**, 131–134.

Qiu, M., Anderson, S., Chen, S., Meneses, J. J., Hevner, R., Kuwana, E., Pedersen, R. A. & Rubenstein, J. L. R. 1996 Mutation of the *Emx-1* homeobox gene disrupts the *corpus callosum*. *Devl Biol.* **178**, 174–178.

Saitou, N. & Nei, M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* **12**, 406–425.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. & Boncinelli, E. 1992 Nested expression domains of four homeobox genes in developing rostral brain. *Nature* **358**, 687–690.

Simeone, A., Acampora, D., Mallamacci, A., Stornaiuolo, A., D'Apice, M. R., Nigro, V. & Boncinelli, E. 1993 Two vertebrate homeobox genes related to the *Drosophila* empty spiracle gene are expressed in the embryonic cerebral cortex. *EMBO J.* **12**, 2735–2747.

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

Takezaki, N., Rzhetsky, A. & Nei, M. 1995 Phylogenetic test of the molecular clock and linearized trees. *Molec. Biol. Evol.* **12**, 823–833.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

Valentine, J. W., Erwin, D. H. & Jablonski, D. 1996 Developmental evolution of metazoan bodyplans: the fossil evidence. *Devl Biol.* **173**, 373–381.

Wray, G. A., Levinton, J. S. & Shapiro, L. H. 1996 Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science* **274**, 568–573.

Received 21 July 1997; accepted 28 August 1997