Saltatory Evolution of the Ectodermal Neural Cortex Gene Family at the Vertebrate Origin

Nathalie Feiner1,2, Yasunori Murakami3, Lisa Breithut1, Sylvie Mazan4, Axel Meyer1,2, and Shigehiro Kuraku1,2,5,*

1Chair for Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Germany
2International Max-Planck Research School (IMPRS) for Organismal Biology, University of Konstanz, Germany
3Department of Biology, Faculty of Science, Ehime University, Matsuyama, Japan
4Développement et Evolution des Vertébrés, UMR7150 CNRS and Université Paris 6, Station Biologique, Roscoff, France
5Present address: Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, Chuo-ku, Kobe, Japan
*Corresponding author: E-mail: shigehiro-kuraku@cdb.riken.jp.

Accepted: July 4, 2013

Data deposition: The molecular sequences identified in this project have been deposited at GenBank under the accession numbers HE981756, HE981757, HE981759, HE981760, and HE981762–HE981764.

Abstract

The ectodermal neural cortex (ENC) gene family, whose members are implicated in neurogenesis, is part of the kelch repeat superfamily. To date, ENC genes have been identified only in osteichthyans, although other kelch repeat-containing genes are prevalent throughout bilaterians. The lack of elaborate molecular phylogenetic analysis with exhaustive taxon sampling has obscured the possible link of the establishment of this gene family with vertebrate novelties. In this study, we identified ENC homologs in diverse vertebrates by means of database mining and polymerase chain reaction screens. Our analysis revealed that the ENC3 ortholog was lost in the basal eutherian lineage through single-gene deletion and that the triplication between ENC1, -2, and -3 occurred early in vertebrate evolution. Including our original data on the catshark and the zebrafish, our comparison revealed high conservation of the pleiotropic expression pattern of ENC1 and shuffling of expression domains between ENC1, -2, and -3. Compared with many other gene families including developmental key regulators, the ENC gene family is unique in that conventional molecular phylogenetic inference could identify no obvious invertebrate ortholog. This suggests a composite nature of the vertebrate-specific gene repertoire, consisting not only of de novo genes introduced at the vertebrate origin but also of long-standing genes with no apparent invertebrate orthologs. Some of the latter, including the ENC gene family, may be too rapidly evolving to provide sufficient phylogenetic signals marking orthology to their invertebrate counterparts. Such gene families that experienced saltatory evolution likely remain to be explored and might also have contributed to phenotypic evolution of vertebrates.

Key words: vertebrate novelty, saltation, gene loss, conserved synteny, whole genome duplication.

Introduction

The first vertebrates emerged more than 500 Ma (Shu et al. 1999; Hedges 2009), and this was paralleled by embryonic novelties, such as the neural crest mainly contributing to craniofacial morphogenesis. The genetic basis underlying these morphological novelties is not fully understood, but increasing sequence data is providing clues to these questions. In particular, recent genome-wide analyses provided convincing evidence of two rounds (2R) of whole-genome duplication (WGD) early in vertebrate evolution (Lundin 1993; Holland et al. 1994; Sidow 1996; Dehal and Boore 2005; Putnam et al. 2008). As a result, the common pattern obtained in phylogenetic analyses of typical gene families is a “four-to-one” relationship in which maximally four vertebrate paralogs are co-orthologs of a single invertebrate proto-ortholog. Among vertebrate lineages, the teleost fishes are characterized by their further derived genomes because of a third round of WGD, the so-called teleost-specific genome duplication (TSGD; Amores et al. 1998; Wittbrodt et al. 1998; reviewed in Meyer and Van de Peer 2005). Postduplication processes, such as neo- or subfunctionalization, based on the initially redundant set of genes, utilized this initial abundance of
genetic raw material for further diversification (Ohno 1970; Force et al. 1999). The redundancy introduced by the 2R-WGD might thus have triggered vertebrate novelties, such as a well-organized brain compartment (Manning and Scheeff 2010).

In addition to the surplus of genomic elements resulting from the 2R-WGD, de novo genes (also often referred to as taxonomically restricted genes or new genes; Khalturin et al. 2009) introduced at the vertebrate origin could have contributed to the vertebrate-specific gene repertoire. A study focusing on genome-wide information of the sea lamprey (Petromyzon marinus), an outgroup to jawed vertebrates, revealed 224 protein-coding genes that are unique to vertebrates (Smith et al. 2013). The target of this study, ectodermal neural cortex (ENC) genes, has been identified only in vertebrates, but they share the conserved BTB/POZ domain and kelch repeats with the rest of the BTB/POZ-kelch repeat superfamily members. The fact that kelch repeat-containing genes are present throughout bilaterians implies that a proto-ENC gene dates back to the last common ancestor of protostomes and deuterostomes (Prag and Adams 2003). The kelch repeat superfamily to which the ENC genes belong is characterized by four to seven tandem repeats of ~50 amino acid motif in a peptide (Bork and Doolittle 1994; Adams et al. 2000). Amino acid sequences between motifs are weakly conserved, except for a few key residues (fig. 1A). This low level of conservation in amino acid sequences impedes a reliable survey of the complete superfamily (Adams et al. 2000).

Despite the divergent amino acid sequence, they all presumably form antiparallel β-sheets that together assemble a β-propeller (Adams et al. 2000). The structural subgroup of the kelch repeat superfamily to which ENC genes belong is additionally characterized by an N-terminal BTB/POZ (Broad-Complex, Tramtrack, and Bric-a-brac/Poxvirus and Zinc-finger) domain of approximately 120 amino acids (Godt et al. 1993; Bardwell and Treisman 1994). This domain is responsible for protein–protein interactions and allows this class of proteins to dimerize (Bardwell and Treisman 1994; Albagli et al. 1995). Proteins encoded by members of the kelch repeat superfamily are implicated in diverse biological processes, and their cellular localizations differ between intracellular compartments, cell surface, and extracellular milieu. Products of several members of this superfamily, including ENC1, have been shown to associate with actin cytoskeleton (Xue and Cooley 1993; Hernandez et al. 1997).

Li et al. (2007) identified ENC1, among others, as suitable phylogenetic marker because it is qualified by the presence of one single coding exon, which facilitates polymerase chain reaction (PCR) amplification with genomic DNA (gDNA). ENC1 as a phylogenetic marker has been employed in numerous phylogenetic studies of actinopterygian fish (e.g., notothenioid fishes [Matschiner et al. 2011], sticklebacks [Kawahara et al. 2009], and ray-finned fishes [Li et al. 2008]) as well as reptiles (iguanian lizards [Townsend et al. 2011] and other squamates [Wiens et al. 2010]).

Hernandez et al. (1997) reported for the first time developmental roles of an ENC gene, namely those of ENC1 in the nervous system of mouse. ENC1 is expressed in a dynamic manner from early gastrulation on throughout neural development and persists in the adult nervous system (Hernandez et al. 1997). A study on various human cell lines suggested that ENC1 is involved in the differentiation of neural crest cells and is down-regulated in neuroblastoma tumors (Hernandez et al. 1998). Interestingly, an antisense transcript of its first exon, ENC1-AS, is linked to a certain type of leukemia (Hammarsund et al. 2004).

Except for mammalian ENC1, only sparse information on the developmental roles of the ENC gene family is available. The expression patterns of chicken ENC1 in the developing telencephalon were characterized in great detail and resemble the dynamic pattern in mouse (Garcia-Calero and Puelles 2009). Expression patterns of the full set of ENC genes (ENC1, -2, and -3) have been investigated only in one species, the amphibian Xenopus laevis (Haigo et al. 2003). The only expression data of ENC genes outside tetrapods are reports of enc3 in developing zebrafish (Kudoh et al. 2001; Thisie B and Thisie C 2004; Thisie C and Thisie B 2005; Bradford et al. 2011; available on the ZFIN database: http://zfin.org/, last accessed July 24, 2013; Qian et al. 2013).

In this study, our exhaustive gene and taxon sampling revealed the diversification pattern of the ENC gene family in a higher resolution. Conserved synteny between genomic regions containing ENC1, -2, and -3 suggested the triplication through 2R-WGDs early in vertebrate evolution. Of those, the ENC3 ortholog was shown to have been lost in the eutherian lineage. We also provide the first report of expression patterns of non-tetrapod ENC1 in a catshark and of the complete set of ENC genes (enc1, -2, and -3) in zebrafish. Overall, molecular and regulatory evolution of the ENC genes within vertebrates conform to typical patterns hitherto observed for many other gene families including developmental regulatory genes, except for one aspect: Conventional molecular phylogenetic methods could not identify the invertebrate orthologs of ENC genes. Because the ENC gene family is one of the numerous subfamilies in the kelch repeat superfamily widely possessed by bilaterians, nonidentification of this long-standing gene in invertebrate indicates unique evolutionary trajectory of the ENC gene family.

Materials and Methods

Collection and Staging of Catshark Embryos

Eggs of the small-spotted catshark Scyliorhinus canicula were harvested by staffs of the Sea Life Centre Konstanz and incubated in separate containers at 18 °C in oxygenated water until they reached required stages. Embryos were dissected in...
FIG. 1.—Comparison of the amino acid sequence of the kelch repeat of selected ENC proteins and phylogenetic relationships within the ENC gene family. (A) The six units of the kelch repeat of all three chicken ENC proteins (ENC1, -2, and -3), the small-spotted catshark ENC1 protein, and all three cyclostome ENC proteins (Eptatretus burgeri ENC-A, Petromyzon marinus ENC-A, and -B) are aligned. Note that the P. marinus ENC-A protein is partial.

(continued)
phosphate-buffered saline solution and staged according to Ballard et al. (1993). Animals that were subjected to in situ hybridizations were fixed for 12 h at 4°C in either Serra's fixative or 4% paraformaldehyde. Additionally, staged and fixed S. canicula embryos were provided by the Biological Marine Resources facility of Roscoff Marine Station in France.

Polymerase Chain Reaction
gDNA extracted from red blood cells of the horn shark Heterodontus francisci and the lemon shark Negaprion brevirostris was gifted by Yuko Ohta. Total RNA was extracted using TRizol (Invitrogen) from a zebrafish at 25 h post-fertilization (hpf), an adult Florida gar Lepisosteus platyrhincus and a S. canicula embryo at stage 33. Total RNA of the inshore hagfish Eptatretus burgeri was gifted by Kinya G. Ota and Shigeru Kuratani. These total RNAs were reverse transcribed into cDNA using SuperScript III (Invitrogen), following the instructions of the 3'-RACE System (Invitrogen).

gDNAs of H. francisci and N. brevirostris, and cDNAs of L. platyrhincus and S. canicula were used as templates for degenerate PCRs using forward oligonucleotide primers that were designed based on amino acid stretches shared among ENC1, -2, and -3 sequences of diverse vertebrates. Forward primer sequences were 5'-GCA TGC WSN MGN TAY TTY GAR GC-3' for the first, and 5'-TGC CAN MGN TAY TTY GAR GCN ATG TT-3' for the nested reaction, and reverse primer sequences were 5'-TG TGC NCC RAA RTA NCC NCC NAC-3' for the first, and 5'-TGC TCC RAA RTA NCC NCC NAC NAC-3' for the nested reaction. The 5'-ends of S. canicula ENC1 and ENC3 transcripts were obtained using the GeneRacer Kit (Invitrogen). These cDNA fragments were used as templates for riboprobes in situ hybridizations.

In addition, the entire 3'-untranslated region (UTR) plus substantial parts of the coding regions of zebrafish enc1, -2, -3, and egr2b (krox20) cDNAs were cloned to prepare riboprobes. Gene-specific primers for these PCRs were designed based on publicly available sequences (ENSDART00000062855 for egr2b, see supplementary table S1, Supplementary Material online, for zebrafish accession IDs). A 249-base pair fragment of E. burgeri ENC-A was identified by performing a TBLASTN search in a hagfish EST archive (http://transcriptome.cdb.riken.go.jp/vtcap/, last accessed July 24, 2013; Takechi et al. 2011) using human ENC1 peptide sequence as query. Based on this sequence, gene-specific primers were designed, and the 5'-part of the coding region plus 5'-UTR of E. burgeri ENC-A was obtained using the GeneRacer Kit (Invitrogen). Assembled full-length S. canicula ENC1 and ENC3 cDNA sequences and the obtained fragments of E. burgeri ENC-A, H. francisci ENC1 and ENC3, N. brevirostris ENC3, and L. platyrhincus ENC2 are deposited in EMBL under accession numbers HE981756, HE981757, HE981759, HE981760, and HE981762–HE981764.

Because the chicken ENC3 gene sequence was incomplete with a stretch of “N”s in the open reading frame (ORF) of ENSGALG00000024263 (Ensemble genome database: http://www.ensembl.org, last accessed July 24, 2013; release 64; Hubbard et al. 2009), we performed a reverse transcriptase (RT)-PCR with gene-specific primers and sequenced the missing part. By aligning the overlapping regions of the deduced protein sequences of the newly obtained fragment and the incomplete sequence in Ensemble, we detected an amino acid substitution. The comparison with other vertebrate ENC proteins clearly showed that this is a highly conserved residue (asparagine). Therefore, we assume that the lysine residue of the Ensemble chicken ENC3 protein was caused by a sequencing error, which is also plausible with respect to the stretch of “N”s. The curated cDNA fragment is deposited in EMBL under accession number HE981758.

Retrieval of Sequences from Public Databases
Sequences of ENC homologs were retrieved from the Ensemble genome database and National Center for Biotechnology Information (NCBI) Protein database, by performing BlastP searches (Altschul et al. 1997) using human ENC1 as query. An optimal multiple alignment of the retrieved ENC amino acid sequences including the query sequence was constructed (fig. 1B) using the alignment editor XCED in which the MAFFT program is implemented (Katoh et al. 2005). Similarly, a second alignment including human, zebrafish, Drosophila

Fig. 1.—Continued
The diagnostic amino acid residues, namely a diglycine followed by a tyrosine, six nonconserved amino acids, and a tryptophan residue are highlighted with gray background. This pattern is disrupted in the first kelch repeat of all three cyclostome proteins where the first glycine (“G”) is replaced by an alanine residue (“A”). Another nonconserved site is a phenylalanine (“F”) instead of a tyrosine (“Y”) in the fourth kelch repeat of the chicken ENC3 protein. Because of similar physicochemical properties, these substitutions do not necessarily prevent the characteristic folding of the mature protein and thus its cellular function. Interestingly, the first kelch repeat of all vertebrate ENC proteins lacks the tryptophan residue and thus does not show the described motif. (B) A phylogenetic tree of the three ENC subgroups of jawed vertebrates, three cyclostome homologs, and the Branchiostoma floridae gene “XP_002612442” as outgroup is shown. Support values are shown for each node in order, bootstrap probabilities in the ML tree inference, and Bayesian posterior probabilities. Analysis is based on 311 amino acids, and the JTT + I + F + Γ4 model was assumed (shape parameter of gamma distribution ¼ 0.66). Red arrows denote sequences that are newly reported in this study. For accession IDs of amino acid sequences used in this analysis, see supplementary table S3, Supplementary Material online.
melanogaster, Ciona intestinalis, and C. savignyi amino acid sequences belonging to the KLHL superfamily was constructed (supplementary fig. S1, Supplementary Material online; for a list of sequences used in this study, see supplementary table S1, Supplementary Material online).

Sea lamprey P. marinus ENC-A was predicted in the AUGUSTUS web server (http://bioinf.uni-greifswald.de/webaugustus/prediction/create, last accessed July 24, 2013) with its species-specific parameters on the supercontig22564 in the version 3 assembly of the genome sequencing project (PMAR3.0). An ORF of the gene designated P. marinus ENC-A was curated (for sequence see supplementary table S2, Supplementary Material online). A truncated fragment of this gene is also present in Ensembl release 64 (ENSPMAG00000008371). The second lamprey ENC gene (ENC-B) is available in Ensembl version 64 (ENSPMAG00000000574). Because of unresolved orthology of these lamprey ENC genes to gnathostome ENC1–3, we refer to them as PmENC-A and PmENC-B.

To search for ENC orthologs in sequenced invertebrate genomes, we explored public databases. Predicted peptide sequences of Nematoestella vectensis, Trichoplax adherens, Helobdella robusta, Capitella teleta, Lottia gigantea, Daphnia pulex, Branchiostoma floridae (all accessible at the DOE Joint Genome Institute: http://www.jgi.doe.gov/, last accessed July 24, 2013) and of Schistosoma mansoni (ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/predictions/, last accessed July 24, 2013) were downloaded, and local Blast searches using human ENC1 protein as query were performed. Invertebrate sequences with high similarity scores were included in the phylogenetic analysis (fig. 2).

Molecular Phylogenetic Analysis

In phylogenetic analyses, we employed PhyML 3.0 (Guindon et al. 2010) for maximum-likelihood (ML) tree inference and MrBayes 3.1 (Huelsenbeck and Ronquist 2001) for Bayesian method. For the ML analyses including large data sets (fig. 2 and supplementary fig. S1, Supplementary Material online), we used RAxML (Stamatakis 2006), because this software tends to outperform PhyML under these conditions (Guindon et al. 2010). Optimal amino acid substitution models were determined by ProtTest (Abascal et al. 2005). To identify invertebrate orthologs of ENC genes and to investigate the phylogenetic relationships within the ENC gene family, a data set that contained relevant representatives of each major vertebrate class for each ENC subtype was created (fig. 1B; see supplementary table S3, Supplementary Material online). We rooted the tree with the most closely related invertebrate protein, B. floridæae XP_002612442 (see below and fig. 2). Similarly, we constructed a molecular phylogeny of the complete KLHL superfamily (supplementary fig. S1, Supplementary Material online). Based on these inferred relationships, several invertebrate sequences that are closely related to the ENC gene family were selected and phylogenetically analyzed for putative orthology to the ENC gene family (fig. 2).

In Situ Hybridization and Immunohistochemistry

The aforementioned 5‘- and 3‘-cDNA fragments of S. canicula ENC1 were used as templates for riboprobes used in situ hybridizations. Paraffin-embedded section in situ hybridizations using S. canicula embryos were performed as described previously (Kuraku et al. 2005), with the modification that the acetylation step and the proteinase K treatment were skipped. Whole-mount in situ hybridizations on catshark embryos were performed according to a protocol originally developed for snake and lizard embryos (Di-Po: N, personal communication). Zebrafish standard whole-mount in situ hybridizations and double in situ hybridizations using the enc1 riboprobes labeled with digoxigenin-UTP and the egrb2 riboprobes labeled with Fluorescein (Roche Applied Science, Mannheim, Germany) were performed as described previously (Begemann et al. 2001; Manousaki et al. 2011). In double in situ staining, enc1 transcripts were detected using nitro blue tetrazolium/5-bromo-4-chloro-3-indoyl-phosphate (BCIP) and egrb2 transcripts by a p-lodontrotetrazolium/BCIP-based detection. Stained embryos were examined with a Zeiss Axioscope microscope. Immunohistochemistry on whole-mount S. canicula embryos was performed as described previously (Kuratan and Eichele 1993) with minor modifications. Monoclonal anti-acetylated tubulin antibody (Sigma T7431) was used to detect developing axons. As secondary antibody, AlexaFluor 568 goat anti-mouse IgG (H + L, Invitrogen A-11004) was applied, and the signal was detected using fluorescence microscopy (Leica). Images were processed with Zeiss Axiosview and Adobe Photoshop software.

Identification of Conserved Synteny

To analyze the mode of the putative loss of ENC3 in eutherians, we downloaded a list of Ensembl IDs of 79 genes harbored in the 1-Mb genomic region flanking ENC3 in chicken, together with IDs of human orthologs of those genes via the BioMart interface. Human orthologs on chromosome 19 were plotted against the corresponding chicken chromosomal region (fig. 3).

We analyzed the genomic regions up to 10 Mb flanking the three chicken ENC genes to search for conserved intragenomic synteny as instructed by Kuraku and Meyer (2012). Using the Ensembl “Gene Tree,” we selected only pairs, triplets, or quartets of paralogous genes that show a gene duplication pattern in accordance with the 2R-WGD (Dehal and Boore 2005). The conserved synteny is depicted in figure 4.

---


1489
FIG. 2.—Phylogenetic tree of vertebrate ENC-related genes of the kelch repeat superfamily and its invertebrate homologs. This tree is based on an alignment of 334 amino acids and was inferred with the ML method assuming the LG + I + F + Γ₄ model (α = 1.67). Support values at nodes are shown in order, bootstrap probabilities in the ML analysis, and Bayesian posterior probabilities. Vertebrate species are color coded in blue, invertebrate deuterostomes in green, and other invertebrates in purple. On the basis of a large-scale phylogenetic analysis encompassing the entire kelch repeat superfamily (supplementary fig. S1, Supplementary Material online), we selected several sequences that are phylogenetically close to the ENC gene family. This selected set of genes was combined with a set of invertebrate homologs that was analyzed for putative orthology to the ENC gene family. Note that the clustering of the Branchiostoma floridae gene “XP_002612442” to the group of ENC genes was only weakly supported by the ML analysis (bootstrap value of 37) and not supported by the Bayesian tree inference.
Results

Identification of ENC Genes in Diverse Nontetrapod Species

By means of RT-PCR, the full-length cDNA of S. canicula ENC1 and ENC3, including 5' and 3'-UTRs, and fragments of E. burgeri ENC-A were sequenced. PCRs using gDNA identified fragments of H. francisci ENC1 and ENC3, N. brevirostris ENC3, and L. platyrhincus ENC2. The inclusion of these genes into the ENC gene family was suggested in BlastX searches in the NCBI nonredundant protein sequence database (nr). These BlastX searches failed to identify any ENC3 orthologs in all available eutherians. An alignment of the deduced amino acid sequences with proteins downloaded from public databases was constructed. The amino acid sequence alignment revealed a high level of conservation especially in the diagnostic residues described previously (fig. 1A; Adams et al. 2000). Each unit of the kelch repeat is characterized by a diglycine followed by a tyrosine, six nonconserved amino acids, and a tryptophan residue (fig. 1A). This pattern is disrupted in the first unit of the kelch repeat of all three cyclostome ENC genes with the first glycine residue replaced by an alanine residue. However, the similar physicochemical property of alanine and glycine theoretically most likely allows this first repeat to be still functional.

Phylogenetic Relationships within Vertebrate ENC

Our sequence data set included selected gnathostome ENC genes and deduced amino acid sequences of the three newly isolated cyclostome ENC genes. Unexpectedly, a protein of a plant, Ipomoea trifida (EU366607 in GenBank), was placed inside the group of teleost ENC1 genes and was found to cluster with stickleback ENC1 (bootstrap support in the ML analysis, 79; data not shown). This placement is in stark contrast to the generally accepted species phylogeny, and therefore we conclude that a contamination of a teleost sequence is the most likely explanation. On the basis of our molecular phylogenetic analysis, we suggest the new gene names enc3 for the formerly called enc1l gene in zebrafish, and Xenc-1 and Xenc-3 for the Xenopus genes previously referred to as Xenc-3 and Xenc-1, respectively (fig. 1B).

1-Mb regions flanking chicken ENC3 (shown in bold) were selected, and gray diagonal lines indicate gene-by-gene orthology between chicken and human. It should be noted that human chromosome 19 is shown in inverted orientation relative to chicken chromosome 28. Human orthologs of the chicken ENC3-neighboring genes, but not ENC3 itself, are concentrated in two distinct regions. The high level of conserved synteny between the chicken ENC3-containing chromosomal region and the human chromosome 19 suggests a small-scale secondary gene loss of ENC3 in the lineage leading to eutherians.
Fig. 4.—Intragenomic conserved synteny between ENC-containing regions in chicken. (A) Overview of the chromosomal location of the three chicken ENC genes (red bars). At the longest, 10-Mb regions flanking the ENC genes were analyzed and are shown in black. The entire region containing paralogs (continued)
The heuristically inferred ML tree (fig. 1B) shows a tight clustering within the three individual subgroups of gnathostome ENC genes (ENC1, -2, and -3). Monophyly of gnathostome sequences for ENC1 (89/0.81), ENC2 (88/1.00), and ENC3 (95/1.00) is inferred (all support values are shown in order, bootstrap probabilities in the ML analysis and Bayesian posterior probabilities; fig. 1B). The three cyclostome ENC genes form an independent group (48/0.65; fig. 1B). The high support (97/1.00) for the clustering of the sea lamprey P. marinus ENC-A with inshore hagfish E. burgeri ENC-A implies their orthology (fig. 1B). The relationship between this cyclostome gene cluster to the three gnathostome ENC subgroups was not unambiguously inferred. The ML tree suggests a closer relationship of gnathostome ENC1 and -3 genes (bootstrap support for their clustering, 27; fig. 1B) to cyclostome ENC genes (bootstrap support, 27; fig. 1B) than to gnathostome ENC2 genes. The topology of the Bayesian analysis inferred a clustering of gnathostome ENC2 and -3 subgroups (posterior probability for their clustering, 0.99; fig. 1B) but did not resolve the trichotomy between this cluster, the ENC1 subgroup, and the group of cyclostome ENC genes. This uncertainty of the phylogenetic position of cyclostome ENC genes demands alternative approaches such as synteny analysis (see below). The exact timings of duplications of the entire genomic region, and thus the ENC gene family, can be pinned down by analyzing the phylogenetic trajectories of neighboring gene families.

Is There an Invertebrate Ortholog of the ENC Gene?
A comprehensive phylogenetic tree was inferred to investigate the relationships of the ENC group of genes to the rest of the KLHL superfamily. This phylogenetic analysis resulted in a close relationship between the vertebrate ENC genes to other genes in the KLHL superfamily, for example, KLHL29 and KLHL30 (supplementary fig. S1, Supplementary Material online). The vast number of sequences was reduced to a data set including only human, zebrafish, D. melanogaster, C. intestinalis, and C. savignyi genes, and a phylogenetic tree was inferred. Based on this comprehensive phylogenetic tree, a subset containing the ENC gene family was selected for further analysis. Sequences of diverse invertebrates were added to this reduced data set, and their position in the tree relative to the ENC gene family was examined (fig. 2). One B. floridiae gene (XP_002612442 in NCBI) was placed close to the ENC group of proteins in the ML analysis (fig. 2). However, this clustering was only weakly supported (bootstrap probability, 37) and was not supported by the Bayesian tree inference (fig. 2). Additionally, a BlastP search of the B. floridiae candidate protein sequence in vertebrates (nonredundant protein sequences in NCBI) revealed its highest similarity to kelch-like protein 24 (KLHL24) instead of the ENC genes. The scaffold57 in the B. floridiae genome assembly (version 1) harboring this B. floridiae gene does not contain any orthologs of the genes surrounding ENC genes in the chicken genome (supplementary table S4, Supplementary Material online). Taken together, our analyses did not particularly support the orthology of this B. floridiae gene (XP_002612442) to the vertebrate ENC genes.

Scale of the Putative Loss of the ENC3 Gene
Our molecular phylogenetic analysis suggested the absence of the ENC3 ortholog in eutherians and possibly in lepidosaurs (fig. 1B). Because of sparse sequence information in the lepidosaurian lineage (genome-wide information only exists for the green anole and the Burmese python [Castoe et al. 2011]), the absence of ENC3 in this taxon is highly speculative at this time point. The absence of ENC3 in eutherians was confirmed by exhaustive TBLastN searches in eutherian genome assemblies using nonmammalian ENC3 peptide sequences as queries. We aimed to determine whether this absence is best explained by a single-gene loss or a large-scale deletion involving substantial parts of the chromosome or even the whole chromosome. For this purpose, we examined whether gene orders are conserved between chicken chromosome 28 containing ENC3 and their orthologs in the human genome. In the region flanking ENC3 (1 Mb both up- and downstream), we identified 62 chicken protein-coding genes that possess orthologs in the human genome, and 58 of these are located on human chromosome 19. More precisely, they are concentrated in two distinct regions (fig. 3). This dense gene-by-gene orthology between these two chromosomes strongly suggests that they are derived from the same ancestral chromosome. Despite several rearrangements, the gene order is well conserved (fig. 3). Thus, a large-scale loss event in the lineage leading to eutherians is not supported. It is more likely that the ENC3 gene was lost in this lineage in a single-gene deletion that did not affect the surrounding genes.

We also attempted to determine the scale of the putative ENC3 loss in lepidosaurs by performing the corresponding analysis between the chicken genomic region containing ENC3 and the orthologous genomic region in the green anole, Anolis carolinensis. However, the orthologs of the
chicken ENC3-neighboring genes were identified on unassembled small contigs. Thus, the current assembly of the *A. carolinensis* genome does not allow us to draw any conclusions about the scale of the putative loss of ENC3.

**Did ENC1, -2, and -3 Arise through the 2R-WGD?**

In addition to the molecular phylogenetic analysis, we addressed the question of the timing of the ENC gene family diversification by investigating the conserved gene order between chicken genomic regions containing ENC1, -2, and -3. The chicken genome was selected for this purpose because it still retains the ENC3 ortholog (unlike eutherians), and it experienced no additional genome duplication (unlike teleosts). The comparisons between the three genomic regions revealed 47 flanking gene families whose pattern of diversification matches the expected 2R-WGD pattern (fig. 4). Additionally, the hypothetical fourth chromosome of the initial 2R-WGD quartet was identified: 15 gene families feature one of the 2R-WGD quartets on chromosome 8 or 25 (fig. 4). The identification of these two chromosomes is not surprising because genome-wide synteny analyses between human and chicken revealed that chicken chromosomes 8 and 25 are orthologous to human chromosome 1 (International Chicken Genome Sequencing Consortium 2004; Voss et al. 2011). This is best explained by chromosome fission in the lineage leading to chicken that gave rise to chromosomes 8 and 25.

**Embryonic Expression Patterns of Catshark ENC1 and Zebrafish ENC1, -2, and -3**

Here, we report the expression patterns of the ENC1 gene in the small-spotted catshark and enc1, -2, and -3 in zebrafish. We performed in situ hybridizations on histological samples of embryos of the small-spotted catshark and whole-mount in situ hybridizations on developing zebrafish. Both 5′- and 3′-riboprobes for the catshark ENC1 gene (see Materials and Methods) yielded the same result, and the expression patterns shown in figure 5 were obtained using riboprobes prepared with the 3′-end cDNAs. Our analysis on catshark embryos at intermediate (stages 26.5–28) and late stages (stages 30–35) of development did not detect any significant expression signal outside the central nervous system (fig. 5). The upregulation was first detected in embryos at stage 26.5, when the expression signal was the most intensified in the corpus cerebelli, the hypothalamus (particularly in the nucleus lobi lateralis), the hindbrain, and a putative sensory patch of the otic vesicle (fig. 5B–E). At stage 30, ENC1 is expressed in the superficial region of the cerebellum, midbrain, and telencephalon (fig. 5G and H). The expression in the telencephalon was primarily restricted to the primordial plexiform layer. ENC1 is expressed in the developing nucleus in the hypothalamus (nucleus lobi lateralis) but not in the neurohypophysis. At stage 33, ENC1 is strongly expressed in a specific layer of the optic tectum (dorsal part of the midbrain), pallium (dorsal part of the telencephalon), and a specific part of the diencephalon (presumably prosomere 2; fig. 5J–L). From this stage on it is evident that ENC1 transcripts in the telencephalon are restricted to the pallium and absent from the subpallium (ventral part of the telencephalon). At stage 35, ENC1 is expressed in the dorsal side of the telencephalon (pars superficialis anterior, pars superficialis posterius, and area peri-ventricularis pallialis) and the choroid plexus, which is the only nonneural expression domain of this gene (fig. 5M and N).

The expression patterns of the three zebrafish *enc* genes shown in figure 6 were obtained with riboprobes spanning the 3′-UTR and substantial parts of the coding region. We found significant expression of all three zebrafish ENC genes (*enc1*, -2, and -3) in developmental stages ranging from 12 to 24 hpf (fig. 6). At early stages of development (14 and 16 hpf; fig. 6A, B, and E), *enc1* transcripts are localized in ventral parts of the forebrain, optic vesicle, distinct parts of the hindbrain, newly formed somites, and the tail bud. The *enc1* expression in the outgrowing tail bud is found in a broad domain of mesenchyme (fig. 6A′). Double stainings with *egr2b*, a marker gene for rhombomeres 3 and 5, revealed that both signals overlap in the hindbrain region. Thus, the *enc1* expression in the hindbrain is also restricted to rhombomeres 3 and 5 (fig. 6C and D). At later developmental stages (24 hpf, fig. 6F and G), the expression of *enc1* in the brain persists but does not extend to the anterior most part of the brain. The tail bud expression is reduced to a small domain of the tip of the tail (fig. 6F). We detected the expression of *enc2* at 12 hpf in anterior parts of the developing brain, distinct parts of the hindbrain, the midline of the posterior trunk, and the tail bud (fig. 6H and I). The expression domain in the hindbrain strongly resembles the expression of *enc1* and is most likely also localized in the rhombomeres 3 and 5 (fig. 6H′ and I′). At 24 hpf, *enc2* transcripts are found in the entire anterior part of the central nervous system and a weak expression signal was detected in the tail bud (fig. 6J). Expression signals of *enc3* at 16 hpf were found in the tail bud and a specific part of the hindbrain (fig. 6K). A dorsal view revealed that the expression in the hindbrain is localized in two lateral structures (fig. 6L). At 24 hpf, expression signal of *enc3* is restricted to specific parts of the hindbrain (fig. 6M).

**Discussion**

The ENC Gene Repertoire in Vertebrates

Our survey in public databases (including databases derived from individual genome sequencing projects), as well as PCR screens, revealed the presence of three ENC subgroups (*ENC1*, -2, and -3) in jawed vertebrates, two ENC genes in the sea lamprey (*ENC-A* and -B), and one in a hagfish (*ENC-A*). An alignment of deduced amino acid sequences of ENC genes revealed a high level of conservation of some key residues
Fig. 5.—Expression patterns of *Scyliorhinus canicula* ENC1 between developmental stages 26 and 35. Panels labeled with letters followed by an apostrophe (’) are magnifications of the corresponding overview picture. (A, F, I) Immunohistochemistry stainings of the neural system (i.e., acetylated tubulin) of *S. canicula* embryos at different developmental stages show overviews of head morphologies. B–E, G, H, and J–N are in situ hybridizations on transverse sections at the levels indicated in A, F, and I. (B–B’) Expression signal in the corpus cerebelli (cocb) and two distinct regions of the diencephalon (di, arrowheads) are shown. (C–C’) ENC1 transcripts are detected in the hindbrain (hb) and the presumptive nucleus lobi lateralis (nlobl) that is part of the hypothalamus (hpt, arrow). (D, D’) Parts of the hindbrain and the anterodorsal lateral line ganglion (alg) are expressing ENC1. (E, E’) Expression signals in the hindbrain are maintained at this level, and expression in a putative sensory patch of the otic vesicle (ov) is detected. (G, G’) ENC1 is expressed in the outermost layer of the midbrain (mb). (H–H’) ENC1 transcripts are located in the corpus cerebelli, the midbrain, and the primordial plexiform layer of the telencephalon (tel). (J–J’) ENC1 transcripts are localized in one specific layer of the optic tectum (ot) and specific regions of the pallium (p). No expression signal was detected (continued)
(fig. 1A). Therefore, we assume that the structure of ENC proteins is conserved among vertebrates.

Our phylogenetic analysis clearly supported the individual clusters of three distinct gnathostome ENC subgroups, namely ENC1, -2, and -3 (fig. 1B). These three subgroups show uniform rates of evolution indicated by comparable branch lengths. Interestingly, we do not detect any additional gene in teleost fish generated in the TSGD (Meyer and Van de Peer 2005). This observation can be best explained through a secondary gene loss of one ENC paralog derived from this third round of WGD before the radiation of teleosts. It is also noteworthy that we did not find any ENC2 gene in multiple chondrichthyan species. Further sequence data of this taxon are needed to confirm a possible loss of chondrichthyan ENC2.

**Origin of the ENC Gene Family**

The ENC gene family is a member of the kelch repeat superfamily (supplementary fig. S1, Supplementary Material online) and shares the conserved BTB/POZ domain and the kelch repeats with other members (fig. 1A). Our database mining and molecular phylogenetic analysis did not identify any apparent ENC ortholog in invertebrates (fig. 2; supplementary table S4, Supplementary Material online). One possible explanation for the alleged absence of invertebrate ENC orthologs might be that they were secondarily lost in invertebrates. However, this assumption would require multiple independent gene losses in diverse invertebrate lineages. Alternatively, this absence can be explained by an elevated evolutionary rate of the ENC gene in the lineage leading to vertebrates erasing significant phylogenetic signals from their sequences (fig. 7). In molecular phylogenies of many gene families, the branch of the lineage leading to vertebrate genes tends to be elongated for the evolutionary time that elapsed for that period. However, the rate of sequence evolution could still be in the range of sufficient gradualism to allow identification of orthology. In contrast, the evolutionary rate of the ENC gene family might have been beyond gradualism, resulting in saltatory sequence change. As a consequence, orthology of vertebrate ENC genes to their counterparts in invertebrates might be no longer traceable with conventional phylogenetic methods based on overall sequence similarity.

We used the *B. floridæ* gene “XP_002612442” to root the tree, although it has not been revealed to be orthologous to vertebrate ENC genes (fig. 1B). However, the placement of a root to the tree allowed us to address the question about the relationship between cyclostome and gnathostome ENC genes. In this study, we identified three ENC homologs of cyclostomes (hagfish and lamprey) that occupy a key phylogenetic position in addressing early vertebrate evolution. In our phylogenetic analysis, the position of the cyclostome ENC genes remains poorly resolved, and no clear orthology to any gnathostome ENC subgroup was confidently suggested (fig. 1B). Depending on the method we applied, alternative scenarios are conceivable, regarding the diversification pattern within the ENC gene family. This unreliability of the molecular phylogeny is enhanced by unclear timing of WGDs (Kuraku et al. 2009). One scenario in which the three jawed vertebrate ENC subgroups originated through gnathostome-specific gene duplications would result in a clustering of all gnathostome ENC genes with the exclusion of cyclostome ENC genes. Our data do not suggest this scenario (fig. 1B). A second possibility based on the 2R-WGD is that the group of cyclostome ENC genes is orthologous to one particular gnathostome ENC subgroup. We did not observe any marked affinity of cyclostome ENC genes to a single gnathostome ENC subgroup. The third possible scenario based on the 2R-WGD is that cyclostomes are the only vertebrate group retaining the fourth ENC subtype, the hypothetical ENC4 gene. This scenario would result in a tree topology inferred by the ML method (fig. 1B), if not only the expected ((A,B),(C,D)) but also a (A,(B,(C,D))) topology is admitted as evidence for a 1-2-4 pattern. Also, the phylogeny inferred by the Bayesian method suggests this scenario (fig. 1B). Thus, our phylogenetic analysis suggests that cyclostome ENC genes are remnants of the fourth ENC subtype that is absent from gnathostome genomes (fig. 7). All scenarios imply an additional cyclostome-specific duplication of the ancestral ENC4 gene resulting in *E. burgeri* ENC-A, *P. marinus* ENC-A and ENC-B followed by a secondary gene loss or nonidentification of the ENC-B gene in hagfish (fig. 7). It was previously proposed that frequent clustering of cyclostome sequences in molecular phylogenetic trees might be caused by a systematic artifact resulting from their unique sequence properties (Qiu et al. 2011). More sequence data of cyclostomes could potentially provide a higher resolution of the ENC gene phylogeny.

**Putative ENC3 Gene Loss in the Eutherian Lineage**

Our molecular phylogenetic analysis suggested the absence of ENC3 genes in eutherians and possibly in lepidosaurs (fig. 1B).

![Fig. 5.—Continued](image-url)

in the epiphysis (epi). (K–K”) Low levels of expression were detected in the corpus cerebelli, whereas strong expression signal was evident in a specific area of the diencephalon, the prosomere 2 (di p2). (L, L’) The ENC1 expression continues more caudally in the hindbrain. (M) The rostral-most part of the pallium, the pars superficialis anterior of the dorsal pallium (pdsa), and the area periventricularis pallialis (app) show ENC1 expression, whereas it is absent from the subpallium (sp). (N) The only nonneural expression domain of ENC1 is the choroid plexus (chp). asb, area superficialis basalis; ed, endolymphatic duct; ob, olfactory bulb; oe, olfactory epithilium; str, stratum; teg: midbrain tegmentum. Scale bars: 0.5 mm in B–E, G, H, and J–N, 100 μm in all magnifications. Smeets et al. 1983 was referred for the morphological identification.
Expression patterns of enc1, -2, and -3 in zebrafish embryos. In situ hybridizations of enc1 (A, B, and E–G), enc2 (H–J), and enc3 (K–M). Expression patterns are shown at 12 hpf (H, I), 14 hpf (A, B), 16 hpf (C–E, K, L), and 24 hpf (F, G, J, M). Panels labeled with letters followed by an apostrophe (‘) are magnifications of the corresponding overview picture. (A–A’, B) Lateral views of enc1 expression reveals signals in ventral parts of the forebrain (arrow), the optic vesicle (opt), distinct parts of the hindbrain (arrowheads), somites (s), and the tail bud (tb) at 14 hpf. (C, D) Lateral view of a double staining (continued)
The secondary loss of the ENC3 gene in the lepidosaur lineage cannot be inferred with high confidence because of sparse sequence information in this lineage. Our attempt to trace conserved synteny between the chicken ENC3-containing genomic region and the green anole genome failed because of insufficient assembly continuity of the latter genome. In contrast, a considerably large number of eutherian genomes have been sequenced, and this speaks in favor of a secondary gene loss instead of incomplete genome sequencing. Other examples of genes that are absent from mammalian genomes, and therefore remained unidentified until recently, include the Bmp16 gene (Feiner et al. 2009), the Edn4 gene (Braasch et al. 2009), the Pdx2 gene (Mulley and Holland 2010), and the Hox14 gene (Powers and Amemiya 2004). To address whether the presumed absence of ENC3 in this lineage was caused by a small-scale secondary loss or rather a large-scale deletion, we searched for conserved synteny between the chicken chromosomal region containing ENC and the human genome. We identified an array of orthologous genes shared between chicken chromosome 28 and human chromosome 19 (fig. 3), as previously suggested by macro-synteny data (International Chicken Genome Sequencing Consortium 2004). The fact that orthologs of chicken ENC3-neighboring genes are present in the human genome suggests a single-gene loss of ENC3 in the common ancestor of eutherians. It is interesting to investigate in future work what impact the loss of the ENC3 ortholog had on associated pathways and to what extent ENC1 and -2 might have possibly compensated the roles of ENC3.

**Expansion of the ENC Gene Family in 2R-WGD**

By performing intragenomic comparison in chicken, we identified a quartet of chromosomes containing ENC1, -2, and -3 and the region that presumably erstwhile harbored the putative fourth paralog (fig. 4). The patterns and timings of duplications in neighboring gene families lend support to the hypothesis that ENC1, -2, and -3 are derived from the 2R-WGD early in vertebrate evolution (Dehal and Boore 2005; Kasahara 2007; Putnam et al. 2008). The precise timing of the 2R-WGD was revealed to be after the split of the invertebrate lineages but before the divergence between cyclostomes and gnathostomes (Kuraku et al. 2009).

**Quartets of chromosomes showing conserved synteny have been used as evidence of the 2R-WGD (Lundin 1993; Holland et al. 1994; Sidow 1996; Spring 1997). It was previously shown that chicken chromosomes 8, 10, 17, 28, W, and Z were derived from one single chromosome in the hypothetical karyotype of the vertebrate ancestor (Nakatani et al. 2007). This set of corresponding chromosomes after the 2R-WGD does not form a quartet but a sextet, possibly...**

---

**Fig. 6.—Continued**

of enc1 and egr2b in a 16 hpf embryo shows overlapping signal in rhombomeres 3 (r3) and 5 (r5). (E–E') Dorsal view of an embryo at 16 hpf reveals enc1 expression in r3 and r5, the tail bud, and additional signal in newly formed somites. (F) Lateral view of expression signal of enc1 in a 24 hpf embryo shows persistence of transcripts in distinct, anterior parts of the brain, and the tail bud. (G) Dorsal view of a 24 hpf embryo indicates that enc1 expression is concentrated in the central nervous system. (H, H') Lateral view of a 12 hpf embryo shows expression in anterior parts of the developing brain (arrow), presumptive r3 and r5, and the tail bud. (I) Dorsal view of the embryo in H reveals additional expression of enc2 along the posterior midline. (J) Dorsal view of a 24 hpf embryo shows enc2 expression in the developing brain and weak expression signal in the tail bud. (K, K') Lateral and dorsal views of enc3 expression signals in a 16 hpf embryo reveals expression in the tail bud and a distinct area of the developing hindbrain (arrowhead). (L) Dorsal view of embryo in K indicates that the hindbrain signal appears in a paired structure. (M, M') Dorsal view at 24 hpf shows enc3 expression in lateral parts of the hindbrain.
because of chromosome fission after the first round of duplication (Nakatani et al. 2007). Our analysis focusing only on parts of the chromosomes harboring ENC genes identified the same set of chromosomes with the exception of chromosome 25, instead of W and 17 (fig. 4). More precisely, our analysis suggested that chromosomes 25 and 8 are derived from one proto-chromosome separated by fission (fig. 4). The incongruence is best explained by different resolution of our study compared with that of Nakatani et al. (2007). Although we focused on a 20-Mb region flanking the ENC genes, the previous study employed fewer markers in the genomic region of our interest (Nakatani et al. 2007). This is why our study provided a higher resolution to detect microlevel genomic rearrangements relevant for ENC gene family evolution (fig. 4).

Conserved Role of ENC Genes in Brain Patterning
Chondrichthyans occupy a key phylogenetic position serving as outgroup to osteichthyes (including teleosts and tetrapods). Comparisons of features between chondrichthyans and osteichthyes allow us to reconstruct the ancestral state of jawed vertebrates. Our study advances the knowledge on both of these major gnathostome lineages by providing the first report of ENC1 expression patterns in a chondrichthyan and expression profiles of all three ENC genes in a teleost. Expression analysis of the full set of ENC genes in a single species was hitherto only performed in the amphibian X. laevis (Haigo et al. 2003). Detailed cross-species comparisons need to be drawn with caution, and only homologous structures of corresponding developmental stages can provide meaningful insights into the evolution of expression patterns and their regulation. In this respect, the expression patterns we obtained in the small-spotted catshark S. canicula and the zebrafish are difficult to compare to - ENC1 -3 because Haigo et al. (2003) mainly focused on earlier developmental stages of X. laevis. In addition, the literature does not contain any detailed description of Xenc expression domains in the developing brain as Garcia-Calero and Puelles (2009) and Hernandez et al. (1997) published for chicken (only telencephalon) and mouse ENC1, respectively. The ENC1 expression in the catshark prosencephalon (primordial plexiform layer of telencephalon and specific parts of the pallium; see fig. 5) has also been described for chicken (García-Calero and Puelles 2009) and mouse (Hernandez et al. 1997). In addition, ENC1 is expressed in diencephalon (hypothalamus and prosomere 2 of the diencephalon), mesencephalon (optic tectum), and rhombencephalon (corpus cerebelli and its caudal extension to the neural tube) of catshark (fig. 5) and mouse (Hernandez et al. 1997). This suggests that the roles of ENC1 in brain patterning were already established in the last common ancestor of chondrichthyans and osteichthyes. Although deep homology between all bilaterian brains has been suggested (reviewed in Hirth and Hirth 2013), integrative centers such as the telencephalon have not been identified in nonvertebrate chordates (Wicht and Lacalli 2005; see also Pani et al. 2012). Thus, well-organized brain structures based on the expansion of the neural tube should be regarded as a vertebrate novelty. Its origin in the earliest phase of vertebrate evolution coincides with the establishment of the ENC gene family involved in brain patterning. It is intriguing to corroborate if the emergence of this gene family contributed to the vertebrate novelty of the tripartite brain.

We also identified differences in expression patterns suggesting lineage-specific changes in developmental programs. ENC1 expression in presomitic mesoderm, the only expression domain outside the nervous system, and dorsal root ganglia of mouse embryos (Hernandez et al. 1997) have not been observed in zebrafish (fig. 6A–G) and Xenopus (Haigo et al. 2003). Vice versa, expression signals of ENC1 in the tail bud of zebrafish (fig. 6A–F) and somites of zebrafish (fig. 6A–E) and Xenopus (Haigo et al. 2003) are absent from the developing mouse (Hernandez et al. 1997). Thus, these expression domains of ENC1 were secondarily modified in the respective lineages. We identified a nonneural expression of ENC1 in the choroid plexus of a catshark embryo at stage 35 (fig. 7N) that has not been identified in any other species to date. The choroid plexus potentially is an ancestral jawed vertebrate ENC1 expression domain that was lost in the lineage leading to osteichthyes or, more parsimoniously, represents an autapomorphic feature of chondrichthyes. The ENC1 expression in the optic vesicle is shared between zebrafish (fig. 6A and B), Xenopus, and mouse but is not observed in catshark embryos (fig. 5) and presumably has been established in the common ancestor of osteichthyes.

Within osteichthyes, expression data of ENC2 and ENC3 genes as well as ENC1 allow inferences of possible shuffling of expression domains. Previously, the full set of ENC1, -2, and -3 genes has been investigated in X. laevis (Haigo et al. 2003), and enc3 expression was analyzed in the zebrafish Danio rerio (Bradford et al. 2011; Qian et al. 2013). Our study describing expression patterns of zebrafish enc1, -2, and -3 combined with a reliable orthology assignment (fig. 1B) allows a solid reconstruction of the evolution of expression domains within osteichthyes. During tailbud stages, all three Xenopus ENC genes are expressed in the neural tube and the otic vesicle, and only ENC1 is expressed in the tail bud. In addition, each gene possesses specific expression domains, such as the dorsal fin, the cement gland, and the pronephric anlage for ENC1 (Xenc-3), ENC2, and ENC3 (Xenc-1), respectively. In comparable stages of zebrafish (~16 hpf), all three enc genes are commonly expressed in the tail bud and the developing brain (fig. 6). Each zebrafish enc gene also has specific expression domains such as somites, midline expression (presumably corresponding to the neural tube), and specific parts of the hindbrain for enc1, -2, and -3, respectively (fig. 6A, I, and K). The comparison of the
overlap between expression domains of individual ENC genes between zebrafish and Xenopus reveals that most likely a different set of genes retained the ancestral expression domains: only Xenopus ENC1, but all three zebrafish enc genes retained expression in the tail bud (fig. 6A, H, and K), and Xenopus ENC1 and -2, but only zebrafish enc1 retained the somite-specific expression domain (fig. 6A). The ENC1 gene is expressed in a more pleiotropic manner than its sister genes ENC2 and ENC3 in zebrafish (fig. 6) and Xenopus (Haigo et al. 2003), suggesting its prevalent role in the developing nervous system. The expression of enc1 and -2 in the rhombomeres 3 and 5 that we observed in zebrafish is absent from Xenopus (Haigo et al. 2003). However, the catshark ENC1 gene also showed expression in the hindbrain (fig. 5B–E and I). Thus, the role of ENC1 in the developing hindbrain might be conserved between chordichthyans and teleosts. Our comparison suggests a shuffling of expression domains among ENC1, -2, and -3 in osteichthians. However, without expression data of ENC2 and -3 in a more basal lineage, for example, chordichthyans, we cannot decide whether losses or gains in the lineages leading to osteichthians or actinopterygians caused these differences in expression profiles. An intriguing question about possible shuffling of ENC expression domains also within tetrapods is currently elusive because of missing ENC2 expression data in mammals and the presumed absence of ENC3 in eutherians. Our expression analysis in the small-spotted catshark S. canicula suggests conserved developmental roles of ENC1 in brain patterning during jawed vertebrate evolution. The comparison of the expression profiles we gained for zebrafish enc1, -2, and -3 genes revealed a differential loss of ancestral expression domains between 2R-derived paralogs.

Perspectives

It is usually the case that we can identify invertebrate orthologs of vertebrate gene families even though they experienced secondary events such as WGDs in the vertebrate lineage. Many of such genes are additional copies of existing genes derived from the WGDs. Otherwise, some genes arose de novo at the base of vertebrate evolution. Interestingly, the ENC family does not belong to these categories, possibly because of the salatory evolution of the ancestral ENC gene early in the vertebrate lineage. This unique feature was masked for a long time by a lack of whole-genome sequences of invertebrates. To our knowledge, Satb1/2 genes (Nechanizky et al. 2012) in the homeobox-containing gene family belong to this category (Burglin and Cassata 2002; Zhong et al. 2008). Our finding renders an insightful theme for future genome-wide studies to reveal more long-standing genes that experienced salatory evolution at the emergence of vertebrates and examine their contribution to phenotypic characters unique to vertebrates.

Supplementary Material

Supplementary tables S1–S4 and figure S1 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Young Scholar Fund of the University of Konstanz and the research grant (KU2669/1-1) from the Deutsche Forschungs gemeinschaft (DFG) to S.K., by the University of Konstanz to A.M., and by the International Max-Planck Research School (IMPRS) for Organismal Biology to N.F. S.M. was supported by the EVOLAX ANR grant (number ANR-09-BLAN-026201) and by Région Bretagne (EVOVERT, grant number 049755). The authors thank Ursula Topel, Adina J. Renz, and Elke Hespeler for technical support in cDNA cloning and in situ hybridization, Dominique Leo for supplying zebrafish embryos, and Sven Tschall, Holger Kraus, and Alexander Dressel from the Sea Life Centre Konstanz for providing catshark embryos. A data set of predicted genes of the lamprey whole-genome sequence was produced by Falk Hildebrand. They are also grateful to Kinya G. Ota and Shigeru Kuratani for providing total RNA of the inshore hagfish and Yuko Ohita for providing gDNA of the horn shark and the lemon shark. Their gratitude extends to two anonymous reviewers for their constructive suggestions.

Literature Cited

Evolution of the ENC Gene Family

Manning G, Scheeff E. 2010. How the vertebrates were made: selective pruning of a double-duplicated genome. BMC Biol. 8:144.


