Gene duplications and evolution of vertebrate voltage-gated sodium channels

Alicia E. Novak[1], Manda C. Jost[2,3], Ying Lu[2], Alison D. Taylor[1],

Harold H. Zakon[2,4] and Angeles B. Ribera[1]

[1]Department of Physiology and Biophysics, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80224
[2]Section of Neurobiology, University of Texas, Austin, TX 78712
[3]Section of Integrative Biology, University of Texas, Austin, TX 78712
[4]The Josephine Bay Paul Center in Comparative and Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, 02543

Corresponding author:

Angeles B. Ribera
Department of Physiology and Biophysics
Mail Stop 8307, RC-1N
University of Colorado at Denver and Health Sciences Center (UCDHSC)
Aurora, CO 80224
Phone: 303 724-4517
FAX: 303 724-4501
E-Mail: angie.ribera@uchsc.edu

Abstract.

Voltage-gated sodium channels underlie action potential generation in excitable tissue. To establish the evolutionary mechanisms that shaped the vertebrate sodium channel □-subunit (*SCNA*) gene family and their encoded Na_v1 proteins, we identified all *SCNA* genes in several teleost species. Molecular cloning revealed that teleosts have eight *SCNA* genes, comparable to the number in another vertebrate lineage, mammals. Prior phylogenetic analyses had indicated that teleosts and tetrapods share four monophyletic groups of *SCNA* genes and that tandem duplications selectively expanded the number of genes in two of the four mammalian groups. However, the number of genes in each group varies between teleosts and tetrapods suggesting different evolutionary histories in the two vertebrate lineages. Our findings from phylogenetic analysis and chromosomal mapping of *Danio rerio* genes indicate that tandem duplications are an unlikely mechanism for generation of the extant teleost *SCNA* genes. Instead, analysis of other closely mapped genes in *D. rerio* supports the hypothesis that a whole genome duplication was involved in expansion of the *SCNA* gene family in teleosts. Interestingly, despite their different evolutionary histories, mRNA analyses demonstrated a conservation of expression patterns for *SCNA* orthologues in teleosts and tetrapods, suggesting functional conservation.

Key words: voltage-gated sodium channel, teleosts, gene families, genome duplication, gene duplication

Introduction

Members of the voltage-gated sodium channel α-subunit family (*SCNA* genes) code for transmembrane Na_v1 proteins that allow the sodium influx required for action potential generation in excitable cells (for review, Catterall et al. 2005). Non-chordate species typically have one to two *SCNA* genes (for review, Goldin 2002). In contrast, molecular characterization of the mammalian *SCNA* gene family has revealed nine genes and an evolutionarily related one coding for a non-voltage-gated channel, Na_x. Each mammalian *SCNA* gene displays a unique expression pattern with respect to tissue specificity and developmental regulation (for review, Goldin 2002). In addition, the encoded Na_v1 channels differ functionally on the basis of channel biophysical properties, toxin binding and pharmacology (Catterall et al. 2005). Despite the extensive information existing about extant mammalian *SCNA* genes, important questions remain regarding their evolutionary relationships with non-mammalian vertebrate orthologues.

Phylogenetic analysis has been previously applied to *SCNA* genes to determine the evolutionary history of the vertebrate *SCNA* gene family (Goldin 2002; Goldin et al. 2000; Lopreato et al. 2001; Piontkivska and Hughes 2003; Plummer and Meisler 1999). The phylogenetic estimates of previous studies differ in significant ways (Fig. 1). For example, results from Plummer and Meisler (1999) and Lopreato et al. (2001) are distinct from other studies in that they both support the hypothesis that two early whole-genome duplications in vertebrates (known as the 2R hypothesis, (e.g., Amores et al. 1998; Ohno 1970; Sidow 1996) led to four *SCNA* genes in the most recent ancestor of teleosts and tetrapods (Teleostomi).

Prior studies have also indicated that, in mammals, three of the four *SCNA*

monophyletic groups are linked to HOX gene clusters, and the fourth is part of a chromosome segment that has been separated from the fourth HOX gene cluster (Plummer and Meisler 1999). In zebrafish, seven *hox* genes exist, each on separate chromosomes and in close proximity to a cassette of associated genes (e.g., distal-less *[dlx]*, sonic hedgehog *[shh]*, frizzled [*fzd]*; Amores et al. 1999). Amores et al. (1998) proposed that the larger number of zebrafish *hox* and associated genes arose during a genome-wide duplication event unique to teleosts (see also Meyer and Schartl 1999; Wittbrodt et al. 1998; Vandepoele et al. 2004). Interestingly, Lopreato et al. (2001) suggested that, in teleosts, this genome-wide duplication event contributed to *SCNA* gene family evolution.

Several considerations now warrant reexamination of *SCNA* gene phylogeny. Since the completion of these studies, additional *SCNA* genes have been isolated in a number of teleost species. Furthermore, prior phylogenetic estimates were either unrooted, or rooted using non-vertebrate gene sequences, the latter increasing the risk of incorrectly reconstructed relationships as a result of long-branch attraction. Further, the existence of a completely sequenced genome and linkage maps for the zebrafish now permits the use of synteny of *scna* with other genes to test hypotheses about evolutionary history. In addition, insights into evolutionary history of a gene family can also be obtained by examination of exon-intron organization (Plummer and Meisler 1999). In mammals, the nine different *SCNA* genes have common exon-intron boundaries in regions that code for transmembrane domains but regions coding for intracellular loops show characteristic group-specific differences between the four different *SCNA* monophyletic groups (Plummer and Meisler 1999).

Here, we identify *SCNA* genes in several teleost species and compare them to known mammalian genes. Then, we use three different approaches to test whether an

additional genome wide duplication in teleosts shaped evolution of the *SCNA* gene family. First, we use a phylogenetic analysis of a diverse taxon sample to compare the histories of gene duplication in teleosts and tetrapods. Next, we determine the chromosomal locations of zebrafish *SCNA (scna*, according to the zebrafish nomenclature convention; http://zfin.org/zf_info/nomen.html*)* genes and test for linkage with HOX genes. Third, we compare the exon organization of specific domains of *D. Rerio scna* genes to that of mammalian *SCNA* genes. In addition, we compare the potential functions of teleost and tetrapod sodium channels by determining expression patterns of newly identified *D. rerio scna* genes. By examining evolutionary relationships and gene expression, our comparative approach provides information about orthology and function of vertebrate *SCNA* genes. Importantly, analysis of two distinct vertebrate lineages allows a rigorous test of the dominant paradigm for the rise of novel gene function (or subfunction) following gene duplication events.

Materials and Methods

Animals

Zebrafish (*D. rerio*) were obtained from a local pet store (Petsmart) and raised at 28.5°C according to established procedures (Westerfield 1995). Gold-lined black knifefish (*Sternopygus macrurus*) and elephant nose mormyrids (*Gnathonemus petersii*) were purchased from tropical fish importers and kept in laboratory tanks. Catfish (*Ictalurus punctatus*) were caught by hook and line near Austin, TX. Lamprey (*Petromyzon marinus*) larvae and adults were provided by the Great Lakes Fishery Commission.

SCNA gene cloning

D. rerio: RNA was collected from zebrafish embryos (10-120 hpf) as well as from isolated adult tissues (brain, spinal cord, eye, cardiac and skeletal muscle) using the RNeasy Kit (Qiagen). An additional proteinase K digestion was performed with cardiac and skeletal muscle samples to remove proteins (e.g., contractile proteins) that interfered with the RNA isolation. Reverse transcription (RT) was performed using the Superscript II (Gibco) protocol. Degenerate primers (IDT) were designed to conserved regions, identified on the basis of mammalian *SCNA* sequence alignments (Table 1). These primers were used for both RT and polymerase chain reaction (PCR) reactions. Nested PCR reactions were performed and PCR products were resolved by gel electrophoresis. Products of interest were cloned into the PCR Script Amp Sk(+) vector (Stratagene) for further analysis, including sequencing.

Portions of the *D. rerio scna* gene sequences were also obtained from The Sanger Center zebrafish genomic database (http://www.ensembl.org/Danio_rerio/). Exon boundaries within genomic *scna* gene fragments were predicted using the

NetGene2 Server (http://www.cbs.dtu.dk/services/NetGene2. PCR reactions were also performed to clone regions identified on the genomic database. Full-length sequences were assembled by combining overlapping cDNA sequences and analyzed using the ClustalW algorithm of the freeware BioEdit v7.05 program (http://www.mbio.ncsu.edu/BioEdit/; Tom Hall, Ibis Therapeutics). Sequences were deposited in GenBank (accession numbers: DQ149503, DQ149504, DQ149505, DQ149506, DQ149507, DQ149508, DQ149509). Recently, *D. rerio scn4aa* and *scn4ab* (DQ22153, DQ221254) were independently cloned by Venkatesh et al., 2005.

***S. macrurus*, *I. punctatus*, *G. petersii* and *P. marinus*:** We sampled *SCNA* genes from a phylogenetically diverse sample of teleost taxa. *SCNA* gene sequences for *Takifugu rubripes* (Percomorpha) were extracted from the Takifugu genome database (http://www.fugu-sg.org/). Within the Ostariophysi, to which *D. rerio* belongs, we sequenced *SCNA* genes from a gymnotiform electric fish (*S. macrurus*) and the channel catfish (*I. punctatus*). We also sequenced *SCNA* genes from a mormyrid electric fish (*G. petersii*) which, as an osteoglossomorph, is a member of the most basal extant teleostean group; and from an agnathan vertebrate (*P. marinus*), which is sister to the gnathostomes, and whose sequences were designated for rooting our tree.

A region coding for Na_v1 transmembrane domains II and III of all *SCNA* genes of several teleost species was cloned. Muscle, brain, heart or electric organ RNA was isolated by homogenization in guanidium thiocyanate (Ausubel et al. 1994) or RNA STAT-60 (Ambion). RT was performed with either random primers or the 3' reverse primer of the subsequent PCR reaction. Reaction parameters were optimized for each species. PCR with degenerate primers resulted in an initial PCR fragment of 370 base pairs (Table 1). PCR products were cloned into the TOPO TA vector (Invitrogen) following the addition of terminal adenines (1 Unit Taq at 72°C, 10 minutes). Inserts

were sequenced and additional rounds of PCR were performed with both specific and degenerate primers to extend sequences of interest. Resultant sequences were up to 2,400 bases in length. Sequence alignments were utilized to insure that each cDNA sequence represented a unique gene. Sequences were deposited in GenBank (accession numbers: *S. macrurus* genes - AF378144, AF378139, DQ286578, AF378143, AF378142, AF378141, AF378140, AY183895; *I. punctatus* genes - AY204538, AY204535, AY204534, AY204533, AY204536, AY204537, AY204532, DQ385608; *G. petersii* genes - DQ275137, DQ275138, DQ275139, DQ275140, DQ275141, DQ275142, DQ275143; *P. marinus genes* - DQ275144, DQ275145).

Genomic DNA isolation and Southern blot analysis

Genomic DNA was isolated from adult wild type *D. rerio* using standard methods (Westerfield 1995). Following digestion with HindIII and EcoRI restriction enzymes, DNA was separated on a 1.5% agarose gel and then transferred to nitrocellulose (Sambrook et al. 1989). The ~200 bp probe, derived from a previously published zebrafish sequence (Tsai et al., 2001), was designed to hybridize to an extremely conserved region of all known mammalian *SCNA* genes (encoding the voltage sensor and part of the pore forming domain - IVS4-S5). To optimize the detection of a single and not multiple bands per *SCNA* gene, the probe design met the following criteria (1) lack of internal HindIII and EcoRI sites, and (2) inclusion within a single exon. Probe template DNA was synthesized by PCR using DNA from *scn8aa* ($zfNa_v1.6$; Tsai et al. 2001) as input and the specific primers F-5' ACTTTGTGTCACCAACATTGTTC and R-5' CTTCACGTAGGCAAAGTTTGAC. The probe was radioactively labeled with ^{32}P-dCTP using the Prime-it Kit (Stratagene). Hybridization was performed at 60°C using standard methods (Sambrook et al. 1989). Filters were washed (1x, 15 min, room

temperature; 3x, 15 min, 37°C; 5 x SSPE, 0.1% SDS) and exposed to X-ray film. Films were developed and digitally scanned. Digital files were imported into Adobe Photoshop.

Sequence alignment and phylogenetic analyses

A matrix of *SCNA* nucleotide sequences coding for Na_v1 transmembrane domains II and III was constructed using coding region sequences from *D. rerio, S. macrurus, I. punctatus, G. petersii*, and *P. marinus*, as well as published sequences for *SCNA* genes from *Homo sapiens* and *Rattus norvegicus*, and matching sequences from the *Gallus gallus*, and *T. rubripes* genome sequence databases. *SCNA* regions coding for transmembrane domains II and III represent the longest region of overlap over which we had confirmed sequence data in all species. *H. sapiens, R. norvegicus, G. gallus,* and *T. rubripes* sequences were acquired from GenBank (http://www.ncbi.nlm.nih.gov/).

Nucleotide sequences were initially aligned using default parameters in ClustalX (Thompson et al. 1994). The alignment was improved manually using MacClade 4.0 (Maddison and Maddison 1992), following predicted amino acid sequences for domains II and III of Na_v1 subunits from *H. sapiens* and *R. norvegicus*. Amino acid sequences and sequence lengths were highly conserved for transmembrane regions, making straightforward the alignment of nucleotides in these conserved regions using known amino acid sequences as a guide. However, the cytoplasmic loop between domain II-III varies extensively in length and sequence between Na_v1 isotypes and therefore we excluded this region from subsequent phylogenetic analyses.

PAUP* 4.0b10 (Swofford 2002) was used to calculate pair wise distances, and to test 24 models of nucleotide evolution using the MrModelblock file from Mr. Modeltest 2.0 (Nylander 2004). The latter program was also used to compare likelihood scores

from the 24 models. For these data, we selected a general time-reversible model with invariable sites and a gamma distribution for variable rate sites (GTR+I+G). Phylogeny was estimated using MrBayes 3.0 (Huelsenbeck and Ronquist 2001), implementing the GTR+I+G model. Four Markov chains of 1,000,000 generations were run at the default temperature (0.2), and every 100th tree was saved to a file. The outgroup was set as *P. marinus.* The burn-in asymptote was determined by plotting tree number against log likelihood. Trees saved before the burn-in asymptote were discarded, and a majority rule consensus of the remaining trees was calculated in PAUP*4.0b10 to estimate posterior probabilities. Four replicates of these Bayesian runs were conducted to insure convergence of the posteriors. The final estimate was rooted using the two Na_v1 sequences from *P. marinus*, which were recovered as sister genes with a posterior value of 100%.

Tree comparisons

We used five different comparison tests to determine whether our estimate of mammalian *SCNA* gene phylogeny explained the molecular data better than previously published alternatives (Fig. 1; Plummer and Meisler 1999; Goldin et al. 2000; Lopreato et al. 2001; Goldin 2002; Piontkivska and Hughes 2003). Because the topology of mammalian *SCNA* phylogeny from Goldin (2002) is identical to that from Plummer and Meisler (1999), these two trees were treated as a single hypothesis. In addition, because the phylogenetic tree from Piontkivska and Hughes (2003) contains a 3-way polytomy, all three possible resolutions of this polytomy were tested as separate hypotheses.

Only human gene sequences were used in the tree comparison tests, because the minimum dataset used in previous studies only included mammalian Na_v1

proteins(Fig.1). We created a NEXUS tree file for each competing topology, and for the topology implied by our Bayesian estimate. We compared the seven trees presented in Figs. 1 and 5. Using the parsimony criterion, we used PAUP 4.0b10 to calculate tree lengths and to conduct Kishino-Hasegawa, Wilcoxon signed-ranks, and winning-sites tests. Using the likelihood criterion, we used PAUP 4.0b10 to calculate likelihood scores, and a parsimony-based estimate of the rate matrix to conduct two-tailed Kishino-Hasegawa tests and one-tailed Shimodaira-Hasegawa tests, both using a RELL bootstrap with 1000 replicates. For all tree comparison tests, we excluded the same poorly conserved regions of data that were excluded from the parsimony and Bayesian phylogenetic analyses, as described above.

Gene mapping

The chromosomal locations of *D. rerio scna* genes were determined using the LN54 radiation hybrid panel (Hukriede et al. 1999) that was generously provided by M. Ekker (Ottawa Health Research Institute, Canada). PCR was performed with gene specific primers designed to amplify an ~200 bp region contained within a single exon. PCR was performed as described (Hukriede et al. 1999) using FastStart Taq DNA Polymerase (Roche Diagnostics) in 96 well plates on a GeneAmp 2700 PCR machine (Applied Biosystems). Products were analyzed by electrophoresis on 0.5% agarose gels. Results were interpreted at http://mgchd1.nichd.nih.gov:8000/zfrh/beta.cgi.

Genomic DNA analysis

Regions of genomic DNA corresponding to *scna* genes of interest were identified using BLAST on the NCBI or Ensembl zebrafish databases. RT-PCR sequences were used to identify the nucleotides corresponding to exons. The number of amino acids coded

for by each exon were counted, and compared between each zebrafish orthologue and rat *SCN5A* (M77235) and *SCN4A* (NM_000334).

Detection of scna mRNAs in adult D. rerio tissue

RT-PCR was performed to determine *scna* mRNA expression in excitable tissues of adult zebrafish. RT was performed with 1 µg of total RNA (brain, spinal cord, eye, cardiac and skeletal muscle) using gene specific reverse primers designed for the Radiation Hybrid experiments. Two rounds of PCR (15 cycles each) were performed using 5 µl of RT product in the first reaction at annealing temperatures empirically determined for each set. Positive control (genomic DNA) and negative (no DNA) controls were routinely included.

Results

Southern blot analysis and molecular cloning indicated that teleosts have eight scna genes

We used standard Southern blot analysis of genomic DNA to provide an initial estimate of the size of the zebrafish *scna* gene family. The probe targeted a region that is highly conserved in all known *SCNA* genes. Seven bands displayed strong hybridization signals (Fig. 2). The probe also recognized an additional band, albeit with a weaker hybridization signal. These data suggested the presence of eight *scna* genes in *D. rerio*.

Next, we isolated *D. rerio scna* genes using a combination of PCR and database mining. Partial sequences for eight different genes were initially isolated using a PCR strategy. To obtain complete open reading frames, we performed additional PCR with specific primers designed on the basis of the partial sequence or information in the Ensembl genomic database (www.ensembl.org). Eight different open reading frames, each corresponding to one of the initial partial sequences, were obtained (Table 2; Supplemental Figure 1, Supplemental Table 1). Despite numerous efforts using PCR, library screening and database mining, no additional *scna* genes were revealed. Thus, both the molecular cloning and Southern analysis results revealed eight *scna* genes in zebrafish.

Independently, partial sequence information was obtained for several other teleost species. PCR analysis of genomic DNA or brain, heart, electric organ, and muscle cDNA identified eight *SCNA* genes each in the weakly electric fish *S. macrurus* and catfish *(I. punctatus*). Similar methods were used to identify seven mormyrid (*G. petersii*) genes. A search of the pufferfish (*T. rubripes*) genome database revealed

eight *SCNA* genes. In sum, the molecular cloning data from several teleost species and Southern blot results of zebrafish genomic DNA revealed eight *SCNA* genes.

Phylogenetic analyses

The total number of unique *SCNA* sequences in our alignment matrix was 70: 8 each from *T. rubripes, D. rerio, I. punctuatus,* and *S. macrurus*; 7 from *G. petersii*; 10 each from *H. sapiens* and *R. norvegicus;* 9 from *G. gallus;* and 2 from *P. marinus*. The Bayesian estimate of phylogeny is given in Fig. 3 with branch lengths in Fig. 4. This tree is rooted with the two *P. marinus* sequences which group here as sister genes (100%). We believe these two sequences represent a unique gene or genome duplication in lampreys, which are also thought to have undergone independent *Hox* cluster duplications (see Fried et al. 2003; Stadler et al. 2004). However, re-rooting our tree with the lamprey genes as sister to one or the other of the two main groups, does not change the relationships between genes in the teleost and tetrapod lineages.

We used phylogenetic analysis to test the evolutionary relationships between *SCNA* genes in teleosts and tetrapods. Our results indicated that, in teleosts, each of the four monophyletic gene groups contained two *SCNA* genes (Fig. 3). In contrast, in mammals, each of the four groups contained one to five genes. In three out of the four groups, a monophyletic group of tetrapod genes was recovered as the sister to a monophyletic group of teleost genes, which is consistent with the hypothesis of independent duplication histories in the two vertebrate lineages. Although our Bayesian estimate did not strongly support a similar topology in the fourth group, comprised of *scn1La* genes and their tetrapod orthologues, a Shimodaira-Hasegawa test using an estimated likelihood rate matrix and estimated gamma distribution could not reject an alternative tree rearranged so that monophyletic teleost *scn1La* genes were sister to

their monophyletic tetrapod orthologues (p=0.21).

Tree comparisons

Our phylogenetic analysis resulted in a double-forked tree (Fig. 5). However, phylogenies from other laboratories have resulted in a "ladder" topology (Fig. 1). To test whether differences in the species used for rooting the tree could account for the differences, we tested the fit of our topology and the other published topologies using lamprey sequences for rooting.

Compared to all of the alternatives we tested, our topology had the shortest tree length and the highest consistency and retention indices (Table 3). On the basis of several criteria and tests, our topology for mammalian *SCNA* gene phylogeny was significantly better at explaining the molecular data than the topology of Goldin et al. (2000) and all three possible resolutions of Piontkivska and Hughes (2003) (Table 3). Our tree was not significantly better at explaining the molecular data than the topologies of Plummer and Meisler (1999), Goldin (2002) or Lopreato et al. (2001) (Table 3).

Under the likelihood criterion, our tree had the best likelihood score. Kishino-Hasegawa and Shimodaira-Hasegawa tests using RELL bootstraps with 1000 replicates found our topology to be significantly better at explaining the data than all three possible resolutions of Piontkivska and Hughes (2003) (Table 3), and the K-H tests also found our topology to be significantly better than Goldin et al. (2000). However, the molecular data could not reject the topology of Plummer and Meisler (1999) and Goldin (2002) or the topology of Lopreato et al. (2001) (Table 3).

Nomenclature for zebrafish scna genes

We assigned names to the zebrafish *scna* genes on the basis of the existing

mammalian nomenclature and guidelines established by the zebrafish community (Goldin et al. 2000; http://zfin.org/zf_info/nomen.html; Table 2). When orthology between a pair of duplicated zebrafish *scna* genes and a single mammalian gene was obvious, assignment of names was straight-forward (e.g., *scn4aa* and *scn4ab* are the co-ortholgues of *SCN4A*; Table 2). However, the four other *scna* genes were equally orthologous to more than one mammalian *SCNA* gene. For these cases, we adopted the zebrafish nomenclature convention and used the designation “like”, abbreviated by the letter "L", to denote similarity, but not orthology (http://zfin.org/zf_info/nomen.html). Further, the numerical identifier in zebrafish (*scn#a*) assumed the lowest number in the mammalian group. For example, the zebrafish co-orthologues of the mammalian group containing *SCN1A*, *SCN2A*, *SCN3A* and *SCN9A* are designated *scn1Laa* and *scn1Lab*. Previously published teleost genes have used an alternative nomenclature that we revise here (Table 2).

D. rerio scna genes map to different chromosomal locations and show linkage with hox clusters

It is thought that the four ancestral *SCNA* genes were each linked to a HOX cluster (Amores et al 1998; Plummer and Meisler 1999). If large-scale genome-wide, rather than select tandem, duplication expanded the teleost *SCNA* family, zebrafish *scna* genes would be expected to (1) map to locations not near each other, and (2) show linkage to a HOX cluster. We took advantage of the genomic databases and mapping strategies available for *D. rerio* to test these predictions.

The eight zebrafish *scna* genes mapped to seven different chromosomes, (*scn1Laa* – 9, *scn1Lab* - 6, *scn4aa* - 3, *scn4ab* - 12, *scn5Laa* - 2, *scn5Lab* - 24, *scn8aa* - 23 and *scn8ab* - 6; Fig. 6). Only two genes, *scn1Lab* and *scn8ab*, mapped to the

same chromosome (6; Fig. 6). However, *scn1Lab* and *scn8ab* mapped to locations that are ~36Mb apart on chromosome 6. These results indicate that a mechanism involving tandem duplications is an unlikely basis for expansion of the *SCNA* gene family in teleosts.

The mapping results were also consistent with the second prediction. All of the eight zebrafish *scna* genes mapped to locations that were syntenically associated with a *hox* gene or genes that belong to a HOX cluster (Fig. 6). For example, although chromosome 6 did not contain a *hox* gene *per se*, some of the genes normally associated with a HOX cluster (e.g., engrailed *(en)*, *dlx*) mapped to chromosome 6. Moreover, the members of a *scna* gene pair (e.g., *scn1Laa* and *scn1Lab*) each mapped near a member of the same duplicated *hox* gene pair (e.g., *hoxaa* and *hoxab*). These results support the hypothesis that a large genome-wide duplication of four ancestral genes was the basis for the expansion of the teleost *SCNA* gene family.

Exon organization of scna genes

We examined the exon organization of regions coding for inter-domain loop regions of zebrafish *scna* genes to test predictions of genome-wide duplication as a mechanism for expansion of the teleost *SCNA* gene family. First, we focused on the inter-domain loop region between transmembrane domains I and II (loop I-II). In mammals, *SCN4A* has two exons coding for loop I-II while genes in the three other *SCNA* groups have four (Dib-Hajj et al. 1999b; George et al. 1993; Plummer et al. 1998; Souslova et al. 1997; Wang et al. 1996). Similar to mammals, one zebrafish gene, *scn4aa*, has two exons, E9 and E10 (Table 3). Interestingly, *scn4ab* has three exons in this region (E9, E10A, E10B). Thus, neither *scn4aa* nor *scn4ab* has four exons coding for the loop I-II region (Table 3), similar to the gene-specific exon organization of mammalian *SCN4A*.

We also examined the regions coding for loop II-III, because it also has variable exon organization in mammals (Dib-Hajj et al. 1999b; George et al. 1993; Plummer et al. 1998; Souslova et al. 1997; Wang et al. 1996). For loop II-III, *SCN5A*, *SCN10A* and *SCN11A* have three exons (E14, E15A, E15B), while other *SCNA* genes typically have only two (E14 and E15 only; (Dib-Hajj et al. 1999b; George et al. 1993; Plummer et al. 1998; Souslova et al. 1997; Wang et al. 1996; Plummer and Meisler 1999). Because the genomic databases were not sufficiently annotated in the relevant regions, we performed RT-PCR analyses. The sizes and sequences of the PCR bands that we cloned indicated that there were only two exons coding for loop II-III in zebrafish *scn5Laa* and *scn5Lab* (Table 3) unlike their mammalian co-orthologues (*SCN5A*, *SCN10A* and *SCN11A)*.

Zebrafish scna genes displayed unique expression patterns

RT-PCR analyses revealed that zebrafish *scn1Laa* and *scn1Lab* were expressed in the central nervous system (brain, spinal cord and eye) but not in skeletal or cardiac muscle. *Scn8aa* and *scn8ab* were also predominantly expressed in neural tissues (brain, spinal cord and eye; Fig. 7). In contrast, *scn4aa, scn4ab, scn5Laa* and *scn5Lab* were expressed in a diverse range of excitable tissues (Fig. 7). For example, s*cn4aa* and *scn4ab* mRNAs were detected in skeletal muscle, brain, spinal cord and eye. In addition, *scn4ab* was expressed in cardiac muscle (Fig. 7). In sum, four zebrafish *scna* genes (*scn1Laa*, *scn1Lab, scn8aa* and *scn8ab*) displayed predominantly neural expression patterns. In contrast, s*cn4aa*, *scn4ab, scn5Laa* and *scn5Lab* transcripts were present in both neural and muscle tissues.

DISCUSSION

Evolution of the SCNA gene family in vertebrates

Previous work indicated that tandem duplications of two of the four ancestral *SCNA* genes resulted in ten *SCNA* genes in mammals with nine coding for Na_v1 channels (Plummer and Meisler 1999). However, how the *SCNA* gene family evolved in other vertebrates had not been resolved. In 2001, Lopreato et al. observed that the phylogenetic relationships emerging for the few previously identified teleost *SCNA* genes were not consistent with tandem duplication of selected ancestors and suggested that genome wide duplication was perhaps involved.

Our data provide independent lines of evidence supporting the hypothesis that a genome wide duplication event played a key role in evolution of the teleost *SCNA* gene family. First, we identified *SCNA* genes in several teleost species and determined their phylogenetic relationships (Figs. 3, 4). The results clearly indicate that teleost *SCNA* genes comprise four sets of duplicated genes. Second, the phylogenetic analyses also support genome wide duplication in teleosts as the basis of expansion of the *SCNA* gene family. Second, we mapped *D. rerio scna* genes and examined their relationship to closely linked genes whose map positions were available on public databases (Fig. 6). The results revealed that the *SCNA* genes do not cluster near each other on a single chromosome, as expected if tandem duplications had occurred. Moreover, each member of a *scna* duplicated pair (e.g., *scna4aa* and *scn4ab*) is chromosomally linked to other duplicated genes (e.g., *hoxba*, *hoxbb*). Further, analysis of genomic structure of the Domain I-II loop of the zebrafish genes supports identification of two *SCN4A* co-orthologues (Table 4).

The hypothesis that a genome-wide duplication event, unique to teleosts, shaped

evolution of the *SCNA* gene family is consistent with the fact that several developmental genes are duplicated in zebrafish (Amores et al. 1998; Ekker et al. 1995; Holland and Williams 1990; Stock et al. 1996; Vandepoele at al. 2004). Interestingly, for many developmental genes, genome wide duplication would predict more genes than are present. For example, whole genome duplication predicts eight *hox* genes, but only seven are present in zebrafish (Amores et al. 1998), suggesting loss of one after the genome wide duplication event. For the seven duplicated *hox* genes that are present, studies of expression suggest that functional roles have been partitioned among duplicates (Jozefowicz et al. 2003; Prince 2002). In contrast to developmental genes, our data indicate that all *scna* duplicates have been retained in *D. rerio*, *T. rubripes*, *I. punctatus* and *S. macrurus*.

The results for *G. petersii* sequences were unique in that this species appears to have orthologues for only one of the 2 teleost genes from each of the 4 monophyletic groups, with independent gene (or genome) duplications in *G. petersii* bringing the total number of genes up to 7 or 8. It is known that osteoglossomorph fishes did not diverge from other teleosts before the third round of genome duplication (Crow et al., 2004; Hoegg et al., 2004; de Souza et al., 2005). The different results for *G. petersii* may be the result of a number of factors, either real (e.g., gene conversion) or artificial (e.g., long branch attraction). However, they do not affect our overall conclusions about gene orthology and different duplication histories in tetrapods and teleosts.

Our parsimony and likelihood based topology tests using an extended data set showed that our phylogenetic estimate is significantly better than that of previous analyses (Table 3). Our results differ most from those of Piontkivska and Hughes (2003; Fig. 1). Interestingly, the tree constructed by Piontkivska and Hughes (2003) nests one *SCNA* gene (*SCN8A*) within a group of genes that reside in a cluster on

another chromosome (*SCN1A*, *SCN2A*, *SCN3A*, *SCN9A*). Their tree also contains a phylogenetic polytomy with the [*SCN1A*, *SCN2A*, *SCN3A*, *SCN9A*] group and a gene (*SCN4A*) on yet another chromosome. In two of the three possible resolutions of the polytomy, *SCN4A* is also nested within the group of five genes residing on a different chromosome. All of the other phylogenies in Fig. 1 are fully resolved trees that are at least consistent with the chromosomal location of the genes.

The two early genome duplications in vertebrates (2R hypothesis; (Lundin 1993; Meyer and Schartl 1999; Sidow 1996) are believed to have occurred close to (one before, one after) the divergence of lampreys and hagfishes. Importantly, re-rooting the topology with the two lamprey genes as sister to one or the other of the two largest clusters does not change our conclusions of Teleostomei gene orthology or gene duplication mechanisms. Based on our findings, it appears that one lamprey *SCNA* gene may have been lost, followed by an independent duplication of the remaining gene, thus giving the sister relationship of the 2 genes we found. This possibility is consistent with the facts that (1) *SCNA* genes are linked to the HOX clusters (Plummer and Meisler 1999) and (2) lampreys are believed to have undergone independent HOX cluster duplications after their divergence from other vertebrates (Fried et al. 2003; Stadler et al. 2004). However, due to the fact that *SCNA* genes are closely affiliated with the HOX clusters, our results do not provide independent support for the 2R hypothesis.

SCNA gene expression in teleosts and mammals

Different evolutionary histories in gene families often result in orthologous genes that have divergent functions. We found that different genomic mechanisms increased the

size of the *SCNA* gene family in mammals and teleosts (tandem duplication in mammals versus whole genome duplication in zebrafish). Our analysis of *snca* expression pattern (Fig. 7), however, raises the possibility that *SCNA* function may have been conserved despite different evolutionary mechanisms the two vertebrate lineages.

For example, *scn8aa* and *scn8ab* are orthologous to the mammalian *SCN8A* gene. In the adult, both zebrafish genes are predominantly expressed in neural tissue (brain, spinal cord and eye), similar to the expression pattern of *SCN8A* in mammals (Krzemien et al. 2000; Schaller et al. 1995; Tzoumaka et al. 2000). Interestingly our RT-PCR analyses also reveal expression in skeletal and cardiac muscle, as recently reported for mammalian *SCN8A* in cardiac tissue (Dhar Malhotra et al. 2001; Maier et al. 2002). Similarly, *scn4aa* and *scn4ab* are orthologous to mammalian *SCN4A*, a gene that is predominantly expressed in skeletal muscle (Trimmer et al. 1989). In zebrafish, we detected abundant expression of *scn4aa* and *scn4ab* in muscle (Fig. 7).

Four zebrafish *scna* duplicates were orthologous to more than one mammalian gene. For example, the mammalian genes *SCN1A*, *SCN2A*, *SCN3A* and *SCN9A* are evolutionarily related to zebrafish *scn1Laa* and *scn1Lab*. Both the mammalian and zebrafish genes in this group display expression patterns that are predominantly neuronal. Similarly, *scn5aa* and *scn5Lab* are orthologous to more than one mammalian gene (i.e., *SCN5A*, *SCN10A*, *SCN11A*). While *SCN5A*, *SCN10A* and *SCN11A* group together phylogenetically, *SCN5A* shows low predicted amino acid identity to *SCN10A* (61.0%) and *SCN11A* (47.3%), suggesting divergent functions. Consistent with this notion, *SCN5A* is primarily expressed in cardiac tissue, while *SCN10A* and *SCN11A* transcripts are found in peripheral nociceptive sensory neurons (Akopian et al. 1999; Dib-Hajj et al. 1999a; Amaya et al. 2000). *SCN5A* expression also occurs in central nervous system regions associated with the limbic system (Donahue et

al. 2000; Hartmann et al. 1999). In zebrafish, both *scn5Laa* and *scn5Lab* are expressed in adult cardiac tissue as well as in neurons, similar to *SCN5A* (Fig. 7). However, neither *scn5Laa* nor *scn5Lab* are detected in sensory neurons during the first five days of development, the embryonic period and tissue type corresponding to the earliest expression of *SCN10A* and *SCN11A* (AE Novak and AB Ribera, unpublished observation). Thus, some of the functions noted for the *SCN5A*, *SCN10A*, *SCN11A* group in mammals do not appear to be present in zebrafish (see also, Sneddon et al. 2003).

Interestingly, we also detected expression of *scn4ab*, *scn5Laa* and *scn5Lab* in tissues in which expression of their mammalian orthologues has not been observed (Fig. 7). Zebrafish *scn4ab* products were amplified from all tissues examined including brain, spinal cord and eye, while *SCN4A* in mammals has only been reported in adult skeletal muscle and heart (Zimmer et al. 2002). Expression of *scn4aa* was also detected in all nervous tissues and skeletal muscle but not cardiac tissue (Fig. 7). Both *scn5Laa* and *scn5Lab* were also expressed in nervous and cardiac tissues. In mammals, *SCN5A* is expressed in cardiac muscle and limited regions of the brain (Hartmann et al. 1999; Rogart et al. 1989), while *SCN10A* and *SCN11A* are expressed in peripheral nociceptive sensory neurons (Akopian et al. 1996; Dib-Hajj et al. 1999a; Sangameswaran et al. 1996). However, mammalian expression of *SCN5A*, *SCN10A* or *SCN11A* has not been reported in spinal cord or eye. Thus, genes that are primarily expressed in muscle types in mammals (*SCN4A* and *SCN5A*) may have different roles in the zebrafish.

Consideration of *SCNA* expression and voltage-gated sodium channels in invertebrate and chordate species provides an interesting perspective for our results (Plummer and Meisler 1999). Invertebrates such as insects and crustaceans, have

sodium-dependent action potentials in neurons, but not muscle (Fatt and Ginsborg 1958; Suzuki and Kano 1977). In contrast, protochordates such as amphioxus have sodium-dependent actions potentials in both neurons and muscle (Hagiwara and Kidokoro 1971; Holland and Williams 1990). Despite these significant physiological differences regarding muscle electrical excitability, proto-chordates are believed to have had the same number of *SCNA* genes as do invertebrates. The proto-chordate may have evolved the ability to express a neural gene in muscle (Hagiwara and Kidokoro 1971; Holland and Williams 1990; Plummer and Meisler 1999). If so, the ancestral vertebrate *SCNA* genes were expressed in both neurons and muscle and the zebrafish *scn4ab*, *scn5Laa* and *scn5Lab* expression patterns may reflect this evolutionary history.

Concluding Remarks

Our results indicate that the final numbers of *SCNA* genes encoding Na_v1 proteins in mammals and teleosts are similar (eight versus nine, respectively). In contrast, invertebrate *SCNA* gene families consist of only one or two members (for review, Goldin 2002). It is possible that the functional complexity of the diverse excitable tissues that have evolved in vertebrates requires multiple *SCNA* genes. Notably, the neural crest, a unique vertebrate specialization, gives rise to the peripheral nervous system where several *SCNA* genes display restricted expression. Further, in invertebrates, *SCNA* genes are only expressed in the nervous system whereas they are additionally expressed in muscle and cardiac tissue in vertebrates. Therefore, the diversity of electrically excitable tissue types in vertebrates may have enhanced selection for multiple yet a similar number of *SCNA* genes despite different evolutionary histories.

Acknowledgments. The authors' work was supported by NIH grants (NS 38937; AEN, ADT and ABR, NS 25513; HHZ and YL and NSF IBN 0236147; MCJ).

References

Akopian AN, Sivilotti L, Wood JN (1996) A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. Nature 379: 257-262.

Akopian AN, Souslova V, England S, Okuse K, Ogata N, Ure J, Smith A, Kerr BJ, McMahon SB, Boyce S, Hill R, Stanfa LC, Dickenson AH, Wood JN (1999) The tetrodotoxin-resistant sodium channel SNS has a specialized function in pain pathways. Nat Neurosci 2: 541-548.

Amaya F, Decosterd I, Samad TA, Plumpton C, Tate S, Mannion RJ, Costigan M, Woolf CJ (2000) Diversity of expression of the sensory neuron-specific TTX-resistant voltage-gated sodium ion channels SNS and SNS2. Mol Cell Neurosci 15: 331-342.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282: 1711-1714.

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K (1994) Current protocols in molecular biology. Wiley, New York.

Catterall WA, Goldin AL, Waxman SG. (2005) Nomenclature and structure-function relationships of voltage-gated sodium channels. Pharmacol Rev 57: 397-409.

Crow KD, Stadler PF, Lynch VT, Amemiya C, Wagner GP (2006) The "fish specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23:121-136.

de Souza FSJ, Bumaschny VF, Low MJ, Rubinstein M (2005) Subfunctionalization of expression and peptide domains following the ancient duplication of the Proopiomelanocortin gene in teleost fishes. Mol Biol Evol 22: 2417-2427.

Dhar Malhotra J, Chen C, Rivolta I, Abriel H, Malhotra R, Mattei LN, Brosius FC, Kass RS, Isom LL (2001) Characterization of sodium channel alpha- and beta-subunits in rat and mouse cardiac myocytes. Circulation 103: 1303-1310.

Dib-Hajj SD, Tyrrell L, Cummins TR, Black JA, Wood PM, Waxman SG (1999a) Two tetrodotoxin-resistant sodium channels in human dorsal root ganglion neurons. FEBS Lett 462: 117-120.

Dib-Hajj SD, Tyrrell L, Escayg A, Wood PM, Meisler MH, Waxman SG (1999b) Coding sequence, genomic organization, and conserved chromosomal localization of the mouse gene *SCN11A* encoding the sodium channel NaN. Genomics 59: 309-318.

Donahue LM, Coates PW, Lee VH, Ippensen DC, Arze SE, Poduslo SE (2000) The cardiac sodium channel mRNA is expressed in the developing and adult rat and human brain. Brain Res 887: 335-343.

Ekker SC, Ungar AR, Greenstein P, von Kessler DP, Porter JA, Moon RT, Beachy PA (1995) Patterning activities of vertebrate hedgehog proteins in the developing eye and brain. Curr Biol 5: 944-955.

Fatt P, Ginsborg BL (1958) The ionic requirements for the production of action potentials in crustacean muscle fibres. J Physiol 142: 516-543.

Fried C, Prohaska SJ, Stadler PF (2003) Independent Hox-cluster duplications in lampreys. J Exp Zoolog B Mol Dev Evol 299: 18-25.

George AL Jr, Iyer GS, Kleinfield R, Kallen RG, Barchi RL (1993) Genomic organization of the human skeletal muscle sodium channel gene. Genomics 15: 598-606.

Goldin AL (2002) Evolution of voltage-gated Na(+) channels. J Exp Biol 205: 575-584.

Goldin AL, Barchi RL, Caldwell JH, Hofmann F, Howe JR, Hunter JC, Kallen RG, Mandel G, Meisler MH, Netter YB, Noda M, Tamkun MM, Waxman SG, Wood JN,

Catterall WA (2000) Nomenclature of voltage-gated sodium channels. Neuron 28: 365-368.

Hagiwara S, Kidokoro Y (1971) Na and Ca components of action potential in amphioxus muscle cells. J Physiol 219: 217-232.

Hartmann HA, Colom LV, Sutherland ML, Noebels JL (1999) Selective localization of cardiac SCN5A sodium channels in limbic regions of rat brain. Nat Neurosci 2: 593-595.

Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59:190-203.

Holland PW, Williams NA (1990) Conservation of engrailed-like homeobox sequences during vertebrate evolution. FEBS Lett 277: 250-252.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, Epstein JA, Barbazuk WB, Li FN, Paw B, Postlethwait JH, Hudson TJ, Zon LI, McPherson JD, Chevrette M, Dawid IB, Johnson SL, Ekker M (1999) Radiation hybrid mapping of the zebrafish genome. Proc Natl Acad Sci USA 96: 9745-9750.

Jozefowicz C, McClintock J, Prince V (2003) The fates of zebrafish Hox gene duplicates. J Struct Funct Genomics 3: 185-194.

Krzemien DM, Schaller KL, Levinson SR, Caldwell JH (2000) Immunolocalization of sodium channel isoform NaCh6 in the nervous system. J Comp Neurol 420: 70-83.

Lopreato GF, Lu Y, Southwell A, Atkinson NS, Hillis DM, Wilcox TP, Zakon HH (2001) Evolution and divergence of sodium channel genes in vertebrates. Proc Natl Acad Sci USA 98: 7588-7592.

Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16: 1-19.

Maddison WP, Maddison DR (1992) MacClade version 3: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

Maier SK, Westenbroek RE, Schenkman KA, Feigl EO, Scheuer T, Catterall WA (2002) An unexpected role for brain-type sodium channels in coupling of cell surface depolarization to contraction in the heart. Proc Natl Acad Sci USA 99: 4073-4078.

Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11: 699-704.

Nylander JAA (2004) MrModeltest (Technical Report). Uppsala University, Evolutionary Biology Centre.

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Berlin.

Piontkivska H, Hughes AL (2003) Evolution of vertebrate voltage-gated ion channel alpha chains by sequential gene duplication. J Mol Evol 56: 277-285.

Plummer NW, Galt J, Jones JM, Burgess DL, Sprunger LK, Kohrman DC, Meisler MH (1998) Exon organization, coding sequence, physical mapping, and polymorphic intragenic markers for the human neuronal sodium channel gene SCN8A. Genomics 54: 287-296.

Plummer NW, Meisler MH (1999) Evolution and diversity of mammalian sodium channel genes. Genomics 57: 323-331.

Prince V (2002) The Hox Paradox: More complex(es) than imagined. Dev Biol 249: 1-15.

Rogart RB, Cribbs LL, Muglia LK, Kephart DD, Kaiser MW (1989) Molecular cloning of a putative tetrodotoxin-resistant rat heart Na+ channel isoform. Proc Natl Acad Sci USA 86: 8170-8174.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Sangameswaran L, Delgado SG, Fish LM, Koch BD, Jakeman LB, Stewart GR, Sze P, Hunter JC, Eglen RM, Herman RC (1996) Structure and function of a novel voltage-gated, tetrodotoxin-resistant sodium channel specific to sensory neurons. J Biol Chem 271: 5953-5956.

Schaller KL, Krzemien DM, Yarowsky PJ, Krueger BK, Caldwell JH (1995) A novel, abundant sodium channel expressed in neurons and glia. J Neurosci 15: 3231-3242.

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6: 715-722.

Sneddon LU, Braithwaite VA, Gentle MJ (2003) Do fishes have nociceptors? Evidence for the evolution of a vertebrate sensory system. Proc Biol Sci 270: 1115-1121.

Souslova VA, Fox M, Wood JN, Akopian AN (1997) Cloning and characterization of a mouse sensory neuron tetrodotoxin-resistant voltage-gated sodium channel gene, Scn10a. Genomics 41: 201-209.

Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP (2004) Evidence for independent Hox gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. Mol Phylogenet Evol 32: 686-694.

Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FH, Weiss KM (1996) The evolution of the vertebrate Dlx gene family. Proc Natl Acad Sci USA **93:** 10858-10863.

Suzuki N, Kano M (1977) Development of action potential in larval muscle fibers in Drosophila melanogaster. J Cell Physiol 93: 383-388.

Swofford DL (2002) PAUP* 4.0: Phylogenetic analysis using Parsimony (* and other methods. Sinauer Associates, Sunderland, Massachusetts.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

Trimmer JS, Cooperman SS, Tomiko SA, Zhou JY, Crean SM, Boyle MB, Kallen RG, Sheng ZH, Barchi RL, Sigworth FJ, Goodman RH, Agnew WS, Mandel G (1989) Primary structure and functional expression of a mammalian skeletal muscle sodium channel. Neuron 3: 33-49.

Tsai CW, Tseng JJ, Lin SC, Chang CY, Wu JL, Horng JF, Tsay HJ (2001) Primary structure and developmental expression of zebrafish sodium channel Na(v)1.6 during neurogenesis. DNA Cell Biol 20: 249-255.

Tzoumaka E, Tischler AC, Sangameswaran L, Eglen RM, Hunter JC, Novakovic SD (2000) Differential distribution of the tetrodotoxin-sensitive rPN4/NaCh6/Scn8a sodium channel in the nervous system. J Neurosci Res *60:* 37-44.

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y (2004) Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA 101: 1638-1643.

Venkatesh B, Lu SQ, Dandona N, See SL, Brenner S, Soong TW (2005) Genetic basis of tetrodotoxin resistance in pufferfishes. Curr Biol 15: 2069-2072.

Wang Q, Li Z, Shen J, Keating MT (1996) Genomic organization of the human SCN5A gene encoding the cardiac sodium channel. Genomics 34: 9-16.

Westerfield M (1995) The Zebrafish book : a guide for the laboratory use of zebrafish (*Brachydanio rerio*). University of Oregon Press, Eugene. Or.

Wittbrodt J, Meyer A, Schartl A (1998) More genes in fish? BioEssays 20: 511-515.

Zimmer T, Bollensdorff C, Haufe V, Birch-Hirschfeld E, Benndorf K (2002) Mouse heart Na^+ channels: primary structure and function of two isoforms and alternatively spliced variants. Am J Physiol Heart Circ Physiol 282: H1007-1017.

Table 1. Degenerate primers used for RT-PCR

Nucleotide Sequence (5'-3')	Targeted Region
D. rerio:	
GGNAARACNTAYATGATHTTYTTYGT	IS6 / Exon 9 (F)
ATGGCNTAYGARGARCARAAYCAG	IS6 / Exon 9 (F)
RTGYTCCATNGCCATRAANAG	IIS1 / Exon 11 (R)
TGGCAYATGAAYGAYTTYTTYCAC	IISS1 / Exon 13 (F)
TAYTTYACHAAYGCHTGGTGYTGG	IIIS3 / Exon 17 (F)
TGGYTHATHTTYWSHATHATGGG	IIIS5 / Exon 19 (F)
GGNTGGATGGAYATHATGTANGC	IIISS2 / Exon 20 (F)
GGRTACATDATRTCCATCCANCC	IIISS2 / Exon 20 (R)
GGNGGNCARGAYATHTTYATGAC	DL III-IV / Exon 22 (F)
GTCATRAADATRTCYTGNCCNCC	DL III-IV / Exon 22 (R)
ATGGTNACHATGATGGTNGARACHG	IVS1 / Exon 23 (F)
TCDGTYTCNACCATCATDGTNACC	IVS1 / Exon 23 (R)
ATHGTNGTNAAYATGTAYATHGC	IVS6 / Exon 24 (F)
TGGGARAARTTYGAYCCNGAYGC	Carobxyl tail / Exon 24 (F)
ACCATDGGCARRTCCATDGC	Carboxyl tail / Exon 24 (R)
S. macrurus*, *I. punctatus*, *G. petersii* and *P. marinus:	
TGCTSGTGTGYYTGATYTTCTGG	IIIS5 / Exon 19 (F)
GTGAAGAAKGAKCCRAAGATGATG	IIIS6 / Exon 21 (R)

Targeted regions are designated in terms of encoded protein domains (e.g. IIISS2, see Table 2). Exon positions are derived from rat *SCN6A* (Plummer et al., 1998). Forward and Reverse primers are noted by "F" and "R", respectively. The following abbreviations label degenerate residues: R=A/G, Y=C/T, K=G/T, S=C/G, H=A/C/T, N=A/C/G/T.

Table 2. Zebrafish *scna* Gene and Na_v1 Protein Nomenclature

Gene Name	Protein Name	Chromosome (Placement, cRay)	Previously cloned teleost genes**
scn1Laa	$Na_v1.1La$	9 (532.65)	Na4
scn1Lab	$Na_v1.1Lb$	6 (175.31)*	Na3
scn4aa	$Na_v1.4a$	12 (184.67)	Na6
scn4ab	$Na_v1.4b$	3 (174.07)	Na1
scn5Laa	$Na_v1.5La$	2 (528.24)	Na2
scn1.5Lab	$Na_v1.5Lb$	24 (473.14)	
scn8aa	$Na_v1.6a$	23 (278.21)	
scn8ab	$Na_v1.6b$	6 (421.08)*	Na5

* *scn1Lab and scn8ab* mapped to the same chromosome. However, the two genes are located 245.8 cRay (~36.4 Mb) apart from each other, suggesting that they did not evolve by tandem duplication of a common ancestral gene.
**Lopreato et al. 2001.

Table 3. Results of parsimony and likelihood based tree comparison tests.

I. Tree Comparisons: Parsimony

Topology[A]	Length	CI	RI	KH	Wilcoxon	winning-sites
1	3345	0.607	0.402	best	best	best
2	3372	0.601	0.398	p=0.0521	p=0.0523	p=0,0575
3	3389	0.599	0.382	p=0.0035*	p=0.0036*	p=0.0061*
4	3348	0.606	0.401	p=0.7099	p=0.7098	p=0.8041
5	3493	0.581	0.335	p<0.0001*	p<0.0001*	p<0.0001*
6	3437	0.590	0.361	p<0.0001*	p<0.0001*	p<0.0001*
7	3491	0.581	0.336	p<0.0001*	p<0.0001*	p<0.0001*

II. Tree Comparisons: Likelihood

Topology[A]	-ln L	2-tailed KH with RELL	SH with RELL
1	15828.40116	best	best
2	15862.58517	p=0.124	0.146
3	15890.51693	p=0.001*	0.352
4	15836.75445	p=0.533	0.465
5	16184.90340	p<0.001*	p<0.001*
6	16024.20430	p<0.001*	p<0.001*
7	16184.22587	p<0.001*	p<0.001*

A, - Toplogy: 1, This study; 2, Plummer and Meisler 1999, Goldin 2002; 3, Goldin et al. 2000; 4, Lopreato et al. 2001; 5, Piontkivska and Hughes, 2003 v1; 6, Piontkivska and Hughes, 2003 v2; 7, Piontkivska and Hughes, 2003 v3. Topology from Plummer and Meisler (1999; 2 in Table) is congruent with that of Goldin (2002; 2, in Table). Three versions of the tree from Piontkivska and Hughes (2003; 5, 6, 7 in Table) were tested, to account for all possible resolutions of a 3-way polytomy. P-values indicated by an asterisk (*) are significant at p<0.05.

Table 4. Exon organization of two cytoplasmic loops of zebrafish *scna* genes

	Loop I-II		Loop II-III	
Gene	**Exons**	**mRNA (bp)**	**Exons**	**mRNA (bp)**
scn1Laa	E9, E10A, E10B, E10C	959	E14, E15	614
scn1Lab	E9, E10A, E10B, E10C	921	E14, E15	683
scn4aa	E9, E10A	485	E14, E15	557
scn4ab	E9, E10A, E10B	419	E14, E15	527
scn5Laa	E9, E10A, E10B, E10C	852	E14, E15	581
scn5Lab	E9, E10A, E10B, E10C	848	E14, E15	638
scn8aa	E9, E10A, E10B, E10C	912	E14, E15	647
scn8ab	E9, E10A, E10B, E10C	921	E14, E15	653

Figure Legends

Fig. 1. Prior hypotheses of mammalian Na_v1 phylogeny. The basic topology of the mammalian Na_v1 family is double-forked tree. The six trees presented are adapted from those of Goldin et al. (2000), Plummer and Meisler (1999), Goldin (2002), Lopreato et al. (2001) and Piontkivska and Hughes (2003).

Fig. 2. Southern blot analysis suggests that the zebrafish *scna* gene family comprises eight genes. Southern blot analysis provided an initial estimate of the number of zebrafish *scna* genes. A “universal” probe, that recognized a conserved sequence in the regions coding for domains IVS4-S5 of all known vertebrate *SCNA* isoforms was used. Arrowheads point to the eight hybridizing bands. Even though teleost and tetrapod genomes have undergone divergent evolutionary pathways, the results indicated that the sizes of the zebrafish and mammalian *SCNA* gene families are similar (eight versus nine, respectively).

Fig. 3. Bayesian estimate of phylogeny for vertebrate voltage-gated sodium channel genes (MCMC x 1,000,000 x 4, GTR + I + G). Tetrapod and teleost sequences are highlighted in blue and green, respectively. The results are consistent with different gene duplication histories in tetrapods (tandem duplications) versus teleosts (whole-genome duplication). Numbers above branches indicate posterior probabilities. GenBank accession numbers are provided for tetrapod sequences. Only GenBank accession numbers appear for chicken sequences, as their nomenclature has not yet been established. Numbers in parentheses following *T. rubripes* entries indicate scaffolds from Assembly Release 3 ("Mayffolds"). For other teleost sequences, see

Materials and Methods. *Scna* gene names given for *G. petersii* and *P. marinus* are temporary because their orthologous relationships are not yet fully established.

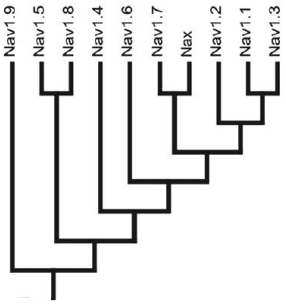
Fig. 4. Bayesian tree with branch lengths. Numbers above branches indicate length. Tetrapod and teleost sequences appear in blue and green text, respectively. GenBank accession numbers appear for chicken sequences, as their nomenclature has not yet been established. Numbers in parentheses following *T. rubripes* entries indicate scaffolds from Assembly Release 3 ("Mayffolds"). See Fig. 3 and Materials and Methods for other GenBank Accession Numbers.

Fig. 5. Vertebrate Na_v1 phylogeny. Our phylogenetic tree for vertebrate Na_v1 proteins is most similar to that previously proposed by Lopreato et al. (2001; Fig.1).

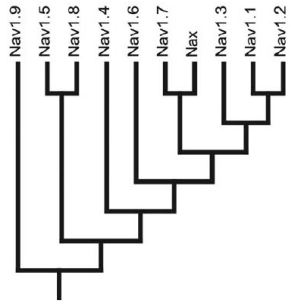
Fig. 6. Each zebrafish *scna* gene mapped to a chromosomal segment associated with a *hox* gene cluster. Radiation hybrid mapping using the LN54 panel revealed that zebrafish *scna* genes (in black) mapped to chromosomes containing HOX clusters. Depicted are paralogous chromosomal segments containing the four sets of duplicated HOX clusters, indicated by Groups A-D. The chromosomal location for each gene is indicated, color-coded by mapping panel: LN54 (cR, red), T51 (cR, blue), HS (cM, green), MOP (icM, orange) and GAT (cM, yellow) (www.zfin.org). In many cases, genes once thought to be found in continuous chromosomal regions have been broken up, and the genes are now found on multiple chromosomes, as has also occurred in humans (Plummer and Meisler 1999). Further, the chromosomal order of the genes shown here has been relaxed to show synteny, as intrachromasomal shuffling has altered the order of many genes. This figure was adapted from Amores et al. 1998.

Fig. 7. RT-PCR revealed novel expression patterns for *scna* genes in electrically excitable tissues of the adult zebrafish. RT-PCR was performed with primers specific for each *scna* gene using RNA isolated from adult tissues. Genomic DNA was used in positive control PCR reactions to assess primer efficacy. Negative controls consisted of RT reactions that lacked the enzyme to detect possible amplification of genomic DNA rather than expressed mRNA.

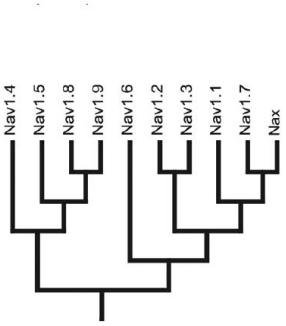
Figure 1:



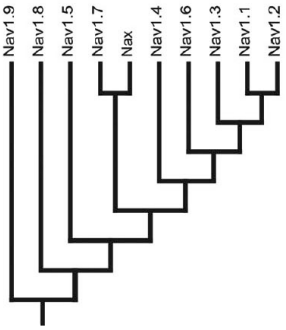
Goldin et al. (2000)



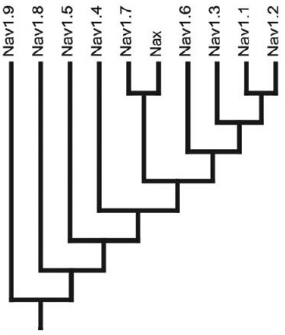
Plummer and Meisler (1999)
Goldin (2002)



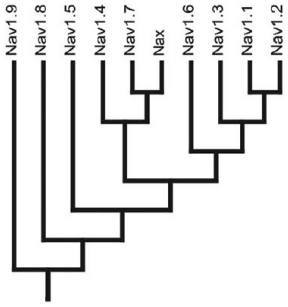
Lopreato et al. (2001)



Piontkivska and Hughes (2003), v1



Piontkivska and Hughes (2003), v2



Piontkivska and Hughes (2003), v3

Figure 2:

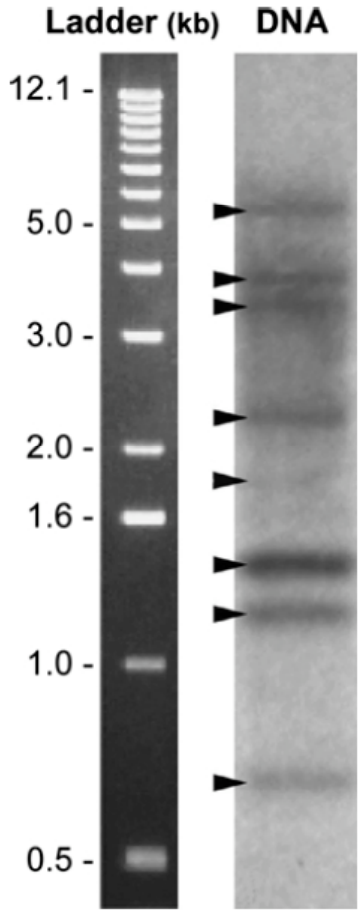


Figure 3:

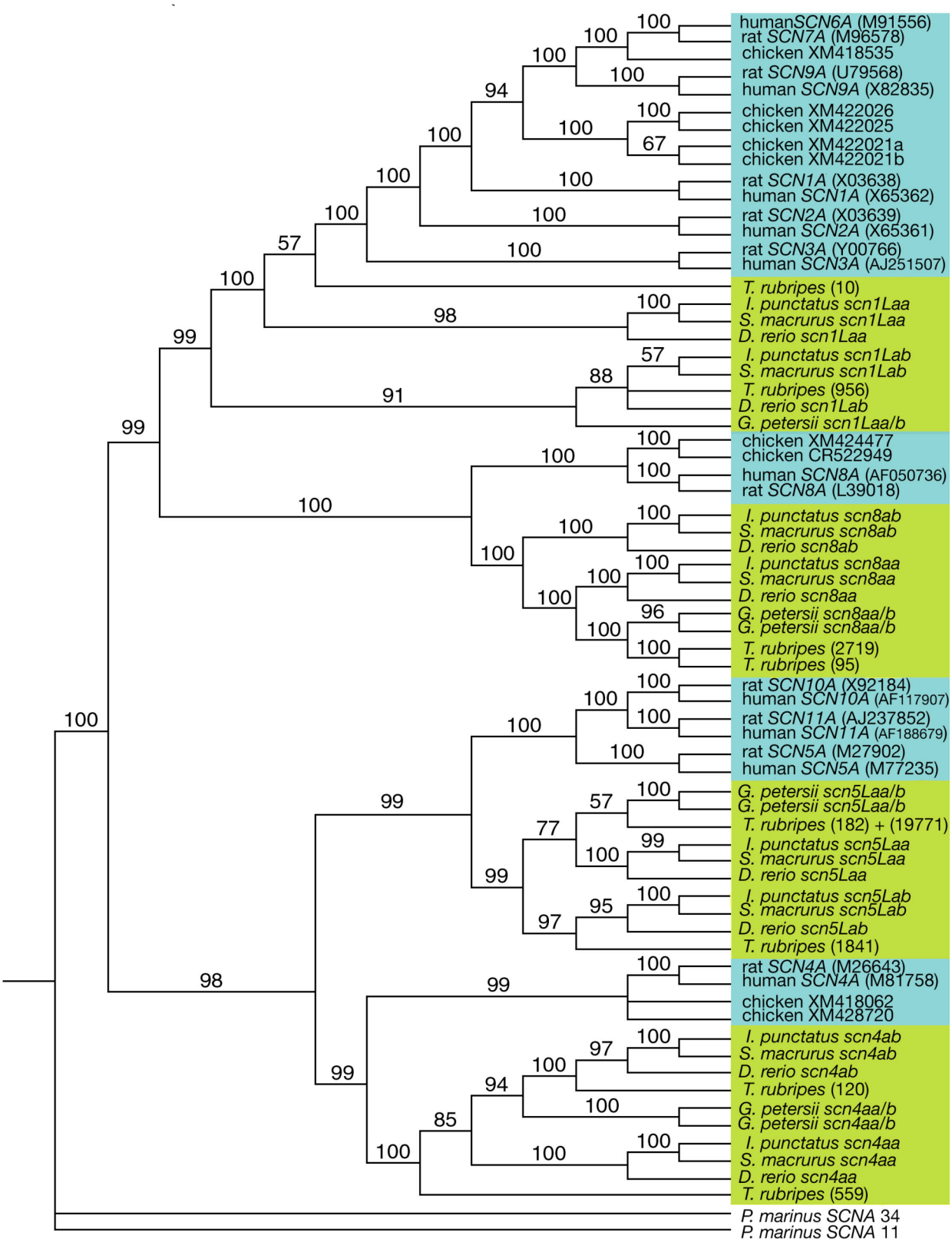

humanSCN6A (M91556)
rat SCN7A (M96578)
chicken XM418535
rat SCN9A (U79568)
human SCN9A (X82835)
chicken XM422026
chicken XM422025
chicken XM422021a
chicken XM422021b
rat SCN1A (X03638)
human SCN1A (X65362)
rat SCN2A (X03639)
human SCN2A (X65361)
rat SCN3A (Y00766)
human SCN3A (AJ251507)
T. rubripes (10)
I. punctatus scn1Laa
S. macrurus scn1Laa
D. rerio scn1Laa
I. punctatus scn1Lab
S. macrurus scn1Lab
T. rubripes (956)
D. rerio scn1Lab
G. petersii scn1Laa/b
chicken XM424477
chicken CR522949
human SCN8A (AF050736)
rat SCN8A (L39018)
I. punctatus scn8ab
S. macrurus scn8ab
D. rerio scn8ab
I. punctatus scn8aa
S. macrurus scn8aa
D. rerio scn8aa
G. petersii scn8aa/b
G. petersii scn8aa/b
T. rubripes (2719)
T. rubripes (95)
rat SCN10A (X92184)
human SCN10A (AF117907)
rat SCN11A (AJ237852)
human SCN11A (AF188679)
rat SCN5A (M27902)
human SCN5A (M77235)
G. petersii scn5Laa/b
G. petersii scn5Laa/b
T. rubripes (182) + (19771)
I. punctatus scn5Laa
S. macrurus scn5Laa
D. rerio scn5Laa
I. punctatus scn5Lab
S. macrurus scn5Lab
D. rerio scn5Lab
T. rubripes (1841)
rat SCN4A (M26643)
human SCN4A (M81758)
chicken XM418062
chicken XM428720
I. punctatus scn4ab
S. macrurus scn4ab
D. rerio scn4ab
T. rubripes (120)
G. petersii scn4aa/b
G. petersii scn4aa/b
I. punctatus scn4aa
S. macrurus scn4aa
D. rerio scn4aa
T. rubripes (559)
P. marinus SCNA 34
P. marinus SCNA 11

Figure 4:

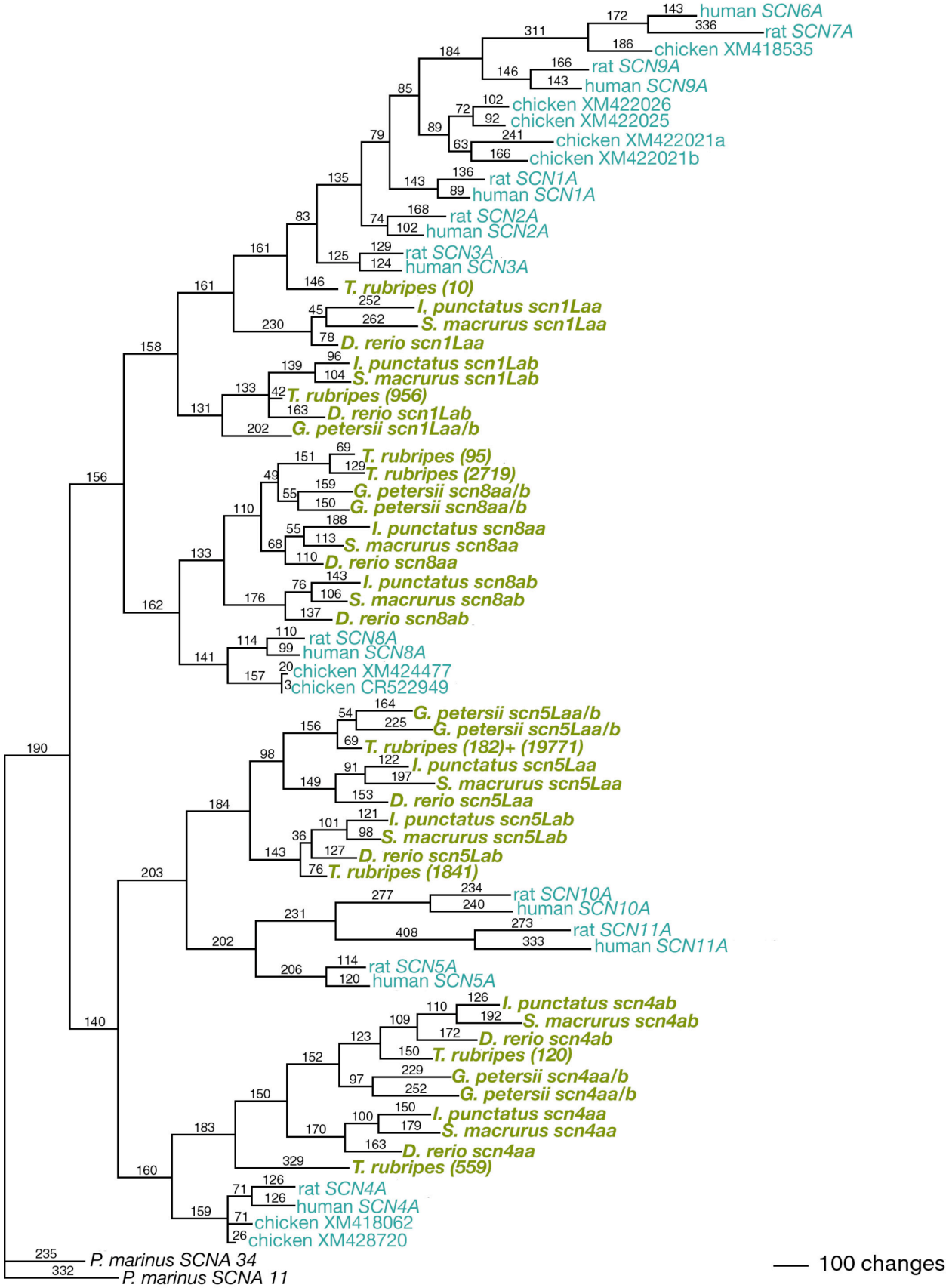

human SCN6A
rat SCN7A
chicken XM418535
rat SCN9A
human SCN9A
chicken XM422026
chicken XM422025
chicken XM422021a
chicken XM422021b
rat SCN1A
human SCN1A
rat SCN2A
human SCN2A
rat SCN3A
human SCN3A
T. rubripes (10)
I. punctatus scn1Laa
S. macrurus scn1Laa
D. rerio scn1Laa
I. punctatus scn1Lab
S. macrurus scn1Lab
T. rubripes (956)
D. rerio scn1Lab
G. petersii scn1Laa/b
T. rubripes (95)
T. rubripes (2719)
G. petersii scn8aa/b
G. petersii scn8aa/b
I. punctatus scn8aa
S. macrurus scn8aa
D. rerio scn8aa
I. punctatus scn8ab
S. macrurus scn8ab
D. rerio scn8ab
rat SCN8A
human SCN8A
chicken XM424477
chicken CR522949
G. petersii scn5Laa/b
G. petersii scn5Laa/b
T. rubripes (182)+ (19771)
I. punctatus scn5Laa
S. macrurus scn5Laa
D. rerio scn5Laa
I. punctatus scn5Lab
S. macrurus scn5Lab
D. rerio scn5Lab
T. rubripes (1841)
rat SCN10A
human SCN10A
rat SCN11A
human SCN11A
rat SCN5A
human SCN5A
I. punctatus scn4ab
S. macrurus scn4ab
D. rerio scn4ab
T. rubripes (120)
G. petersii scn4aa/b
G. petersii scn4aa/b
I. punctatus scn4aa
S. macrurus scn4aa
D. rerio scn4aa
T. rubripes (559)
rat SCN4A
human SCN4A
chicken XM418062
chicken XM428720
P. marinus SCNA 34
P. marinus SCNA 11
100 changes

Figure 5:

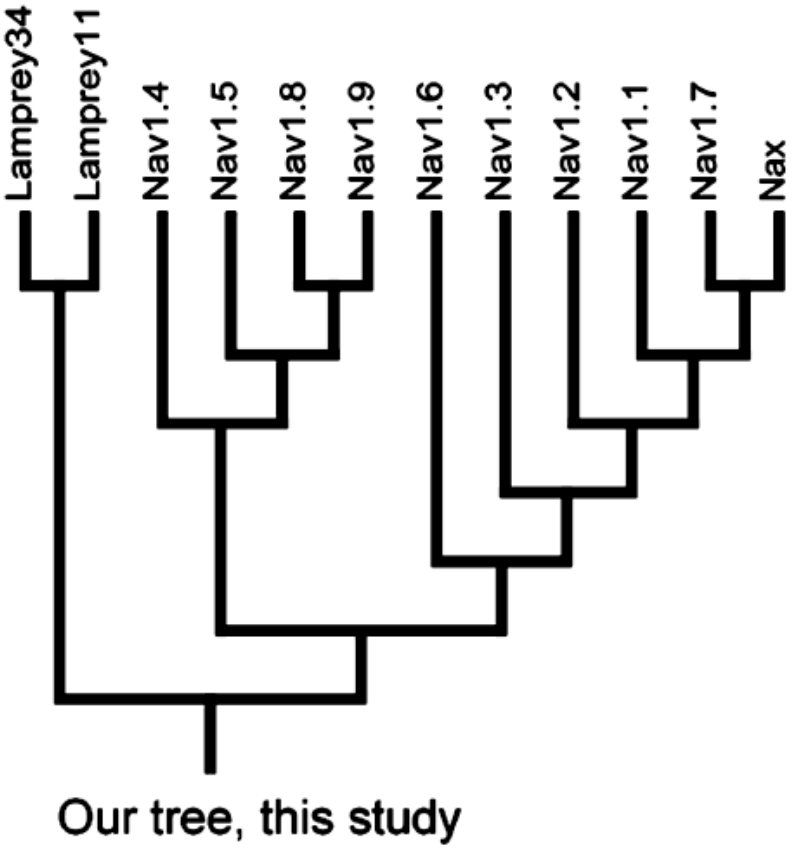

Lamprey34
Lamprey11
Nav1.4
Nav1.5
Nav1.8
Nav1.9
Nav1.6
Nav1.3
Nav1.2
Nav1.1
Nav1.7
Nax
Our tree, this study

Figure 6:

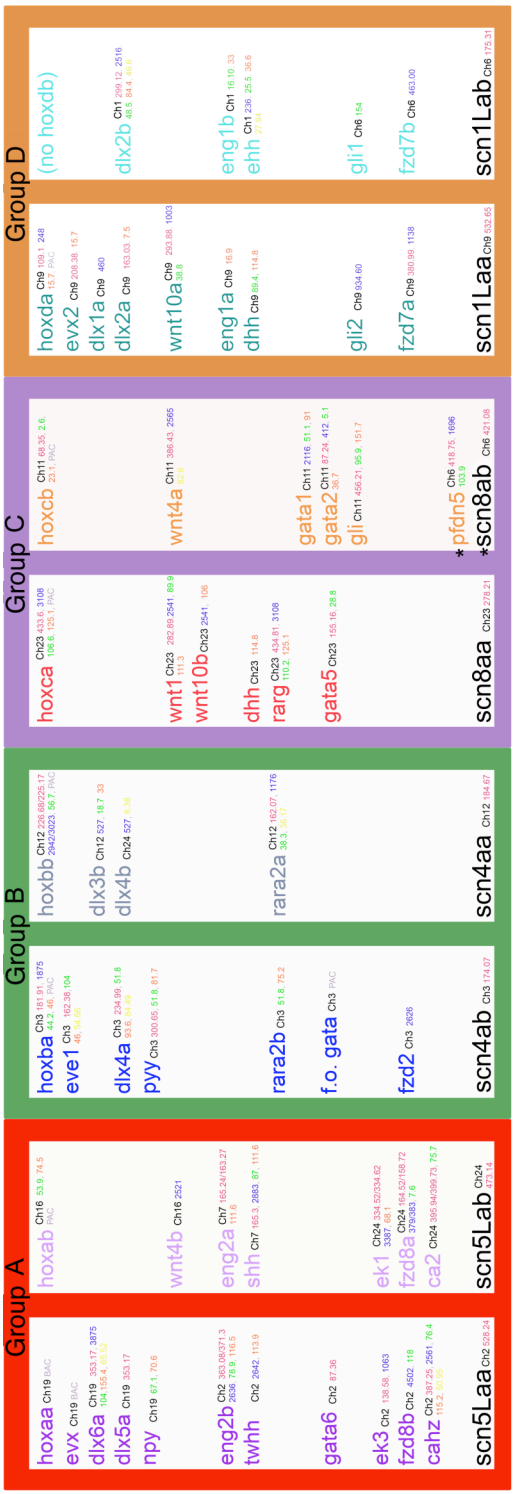
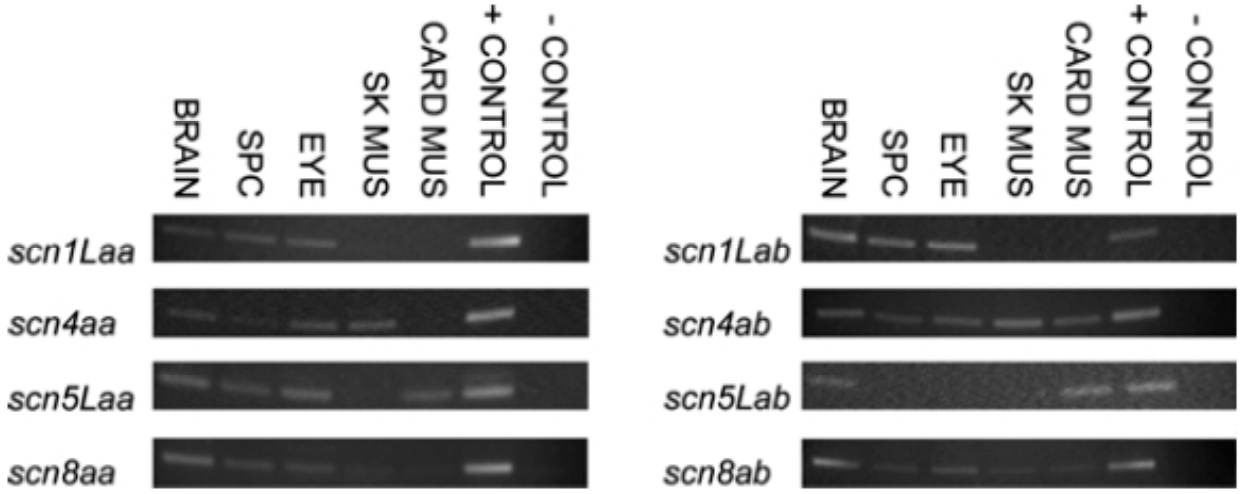

Group A
Group B
Group C
Group D
hoxaa
evx
dlx6a
dlx5a
npy
eng2b
twhh
gata6
ek3
fzd8b
cahz
scn5Laa
hoxab
eng2a
shh
ek1
fzd8a
ca2
scn5Lab
hoxba
eve1
dlx4a
pyy
rara2b
f.o. gata
fzd2
scn4ab
hoxbb
dlx3b
dlx4b
rara2a
scn4aa
hoxca
wnt1
wnt10b
dhh
rarg
gata5
scn8aa
hoxcb
wnt4a
gata1
gata2
gli
*pfdn5
*scn8ab
(no hoxdb)
hoxda
evx2
dlx1a
dlx2a
dlx2b
wnt10a
eng1a
dhh
eng1b
ehh
gli2
gli1
fzd7a
fzd7b
scn1Laa
scn1Lab

Figure 7:


BRAIN
SPC
EYE
SK MUS
CARD MUS
+ CONTROL
- CONTROL
scn1Laa
scn4aa
scn5Laa
scn8aa
BRAIN
SPC
EYE
SK MUS
CARD MUS
+ CONTROL
- CONTROL
scn1Lab
scn4ab
scn5Lab
scn8ab