Evolution of Vertebrate Genes Related to Prion and Shadoo Proteins—Clues from Comparative Genomic Analysis

Marko Premžl,* Jill E. Gready,* Lars S. Jermiin,† Tatjana Simonic,‡ and Jennifer A. Marshall Graves§

*Computational Proteomics Group, John Curtin School of Medical Research, Australian National University, Canberra, Australia; †School of Biological Sciences and Biological Informatics & Technology Centre, University of Sydney, NSW, Australia; §Dipartimento di Patologia Animale, Igiene e Sanità Pubblica Veterinaria, Sezione di Biochimica e Fisiologia Veterinaria, Università di Milano, Milan, Italy; and ‡Comparative Genomics Group, Research School of Biological Sciences, Australian National University, Canberra, Australia

Recent findings of new genes in fish related to the prion protein (PrP) gene PRNP, including our recent report of SPRN coding for Shadoo (Sho) protein found also in mammals, raise issues of their function and evolution. Here we report additional novel fish genes found in public databases, including a duplicated SPRN gene, SPRNB, in Fugu, Tetraodon, carp, and zebrafish encoding the Sho2 protein, and we use comparative genomic analysis to analyze the evolutionary relationships and to infer evolutionary trajectories of the complete data set. Phylogenetic footprinting performed on aligned human, mouse, and Fugu SPRN genes to define candidate regulatory promoter regions, detected 16 conserved motifs, three of which are known transcription factor–binding sites for a receptor and transcription factors specific to or associated with expression in brain. This result and other homology-based (VISTA global genomic alignment; protein sequence alignment and phylogenetics) and context-dependent (genomic context; relative gene order and orientation) criteria indicate fish and mammalian SPRN genes are orthologous and suggest a strongly conserved basic function in brain. Whereas tetrapod PRPNs share context with the analogous stPrP-2–coding gene in fish, their sequences are diverged, suggesting that the tetrapod and fish genes are likely to have significantly different functions. Phylogenetic analysis predicts the SPRN/SPRNB duplication occurred before divergence of fish from tetrapods, whereas that of stPrP-1 and stPrP-2 occurred in fish. Whereas Sho appears to have a conserved function in vertebrate brain, PrP seems to have an adaptive role fine-tuned in a lineage-specific fashion. An evolutionary model consistent with our findings and literature knowledge is proposed that has an ancestral prevertebrate SPRN-like gene leading to all vertebrate PrP-related and Sho-related genes. This provides a new framework for exploring the evolution of this unusual family of proteins and for searching for members in other fish branches and intermediate vertebrate groups.

Introduction

The mammalian PRNP gene (Lee et al. 1999) encodes prion protein (PrP) whose aberrant metabolism is a hallmark of prion diseases (Prusiner 1998). Physiologically, PrP exhibits environment-dependent conformational flexibility, with one isoform, PrP\(^V\), undergoing refolding into a compact, amyloidogenic isoform, PrP\(^Sc\), during prion disease pathogenesis. Both isoforms are encoded by the same PRNP gene. PrP mutations determine disposition or resistance to human familial and infectious forms of prion diseases and determine their phenotype in human, mouse, and sheep (Prusiner and Scott 1997). The normal functions of PrPs are still unknown, although evidence for a bewildering number of functions, often contradictory, has been presented (Aguzzi and Hardt 2003).

The first gene distal to PRNP in eutherian mammals is PRND, which encodes the doppel (Dpl) protein (Moore et al. 1999) and is thought to have arisen by a duplication of PRNP (Mastrangelo and Westaway 2001). The next neighbor gene to PRND, detected so far only in mammals and not present in mouse, is the related PRNT gene (Makrinou, Collinge, and Antoniou 2002), which seems to be a pseudogene that has arisen from duplication of PRND.

The complexity of the mammalian PRNP locus and the conundrum of PrP function provided the impetus for searching publicly available genomic sequence databases of human (Lander et al. 2001) and model vertebrates (Aparicio et al. 2002; Waterston et al. 2002) to discover PrP paralogs in silico. As these paralog genes may have functions related to those of PrP, their characterization may illuminate its functions and reveal the basis for PrP’s disease associations. The paralogs themselves may be indicated to be potential drug targets (Lander et al. 2001).

Using this comparative genomics approach, we recently manually annotated a new gene, SPRN, that is a paralog of PRNP (Premžl et al. 2003). SPRN was detected in mammals (human, mouse, rat) and fish (Fugu, Tetraodon, zebrafish). It encodes the Shadoo (Sho) protein, and is better conserved between mammals and fish than is PRNP (fig. 1).

Several other genes encoding proteins with homology to PrP have also been described recently in fish only (table 1): PrP-like protein from Fugu, Tetraodon, and zebrafish (Suzuki et al. 2002), a protein from Fugu and Tetraodon named PrP461 (Rivera-Milla, Stuermer, and Malaga-Trillo 2003) or stPrP-1 (Oidtmann et al. 2003), and stPrP-2 from Fugu and stPrP from salmon (Oidtmann et al. 2003). In contrast to Shos, these new fish proteins are not well conserved among the major fish lineages nor among mammalian or other vertebrate PrPs (frog, reptile, and bird).

However, despite divergence of their amino acid sequences, many overall protein features peculiar to these proteins are conserved; these features are summarized in figure 1. First, all these proteins are extracellular and predicted to be GPI-anchored and, with the exception of the most divergent PrP-like proteins, are also actually or
potentially N-glycosylated. In addition, PrPs and stPrPs do, or can be predicted to, contain a disulfide bond. Next, all mature proteins have basic regions and, except for frog PrP, repeats. The most characteristic feature of all PrPs and fish proteins related to PrP is a middle stretch of hydrophobic amino acids, which in mammals is essential for the pathogenic PrP isotype conversion (Prusiner 1998). Sho protein shares with these PrPs this unique property of a middle non–membrane-spanning hydrophobic region (fig. 1). In contrast, the Dpl sequence lacks both repeat and hydrophobic

![Diagram showing overall structures of prion proteins (PrPs), PrP-related proteins from fish (stPrP-2, stPrP-1, stPrP-3, PrP-like protein) and Shadoo (Sho) proteins.](https://academic.oup.com/mbe/article-abstract/21/12/2210/1071063)

**Figure 1.**—Overall structures of prion proteins (PrPs), PrP-related proteins from fish (stPrP-2, stPrP-1, stPrP-3, PrP-like protein) and Shadoo (Sho) proteins. Numbers indicate the first residue of each section and last one of each protein. S, signal sequence; B, basic region; R, repeats or low complexity region; PGH, Pro-Gly-His rich; RG, Arg-Gly rich; H, hydrophobic region; N, N-glycosylation site; S-S, disulphide bond; GPI, glycosylphosphatidylinositol anchor. (A, F) Numbers refer to human; (E) numbers refer to Fugu; (G, H), numbers refer to zebrafish.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of Vertebrate Proteins and Genes Related to PrP, Dpl, and Sho</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Protein</td>
</tr>
<tr>
<td>PrP-like</td>
<td>PrP-like</td>
</tr>
<tr>
<td>stPrP-1</td>
<td>PrP461 or stPrP-1</td>
</tr>
<tr>
<td>stPrP-2</td>
<td>stPrP-2</td>
</tr>
<tr>
<td>stPrP-3</td>
<td>stPrP</td>
</tr>
<tr>
<td>SPRN</td>
<td>Sho</td>
</tr>
<tr>
<td>SPRNB</td>
<td>Sho2</td>
</tr>
<tr>
<td>PRNP</td>
<td>PrP</td>
</tr>
<tr>
<td>PRND</td>
<td>Dpl</td>
</tr>
<tr>
<td>PRNT</td>
<td>PRNT</td>
</tr>
</tbody>
</table>
regions and shows homology only to the C-terminal folded domain of PrP.

In addition to the serious problem of transmission of prion disease among mammals and to humans in the recent bovine spongiform encephalopathy (BSE) crisis arising from feeding of contaminated mammal-derived proteins to cows, these new findings of PrP homologs in fish raise new issues: the possibility of spread of prion disease to farmed fish from mammals (and vice versa?) from meat and bone meal feedstuff derived from farm animals. Hence, understanding the evolution of the fish proteins and definition of their functions appear both critical and timely for defining the function, and dysfunction, of the mammalian PrPs, as a means for understanding the transmissibility of infectious prion disease.

Genes in different species defined as orthologous by their derivation from a common ancestor generally lie in the same genomic context and perform the same function; this property is the basis for prediction of gene function in newly sequenced genomes (O’Brien et al. 1999; Waterston et al. 2002). Although this comparative genomics approach is a powerful tool for understanding biological function, the appropriate evolutionary distance of species in a comparison depends on the biological question addressed (Frazer et al. 2003). For example, a comparative study of custom-sequenced genomic data for four yeast species that diverged 5 to 20 MYA allowed a complete analysis for gene identification, determination of gene structures, estimate of rapid/slow evolutionary changes, detection of regulatory elements, detection of combinatorial control of expression, and identification of the key regulatory elements determining gene activity (Kellis et al. 2003). However, there are limitations in the depth and extent of possible mammalian comparative genomic analysis, which parallel availability of eutherian, marsupial, and monotreme lineage representatives (O’Brien, Eizirik, and Murphy 2001; Graves and Westerman 2002). The most comprehensive targeted comparative genomic analysis reported to date included 13 vertebrate species (one human, eight other mammals, one bird, and three fish) and allowed identification of conserved coding and noncoding sequences as well as insight into the genome dynamics (Thomas et al. 2003).

The aim of this work is to analyze the evolutionary relationships among mammalian PRNP and fish genes encoding proteins related to PrP, and also their relationships with SPRN genes from mammals and fish, by means of comparative genomic analysis. This analysis comprises both homology and nonhomology criteria for assessing gene orthology (Eisen and Wu 2002). Apart from identification of gene similarity (homology criteria), we tested whether homologous rearrangement events have occurred in the intergenic regions (nonhomology criteria). Conserved order and transcription orientation with respect to adjacent gene or genes, also dubbed “conserved contiguity,” implies that there has been no rearrangement in the intergenic regions and that the genes share common evolutionary history. This is suggested to be a strong indication of orthology (Comparative Genome Organization Workshop 1996; Gilligan, Brenner, and Venkatesh 2002).

Our analysis is based on sequence data available in public databases. Apart from human, only two vertebrate genome sequences are available in draft form: those for mouse (Waterston et al. 2002) and tiger pufferfish Fugu rubripes (Aparicio et al. 2002). Substantial data for the rat genome is also available (http://www.ensembl.org/Rattus_norvegicus/). Rodents are separated from human by 75 Myr of independent evolution, and mouse and rat (separated by 30 Myr) are the major mammalian models for genetic and immunological research and physiological research, respectively. Data from the recently reported dog genome, which is at only 1.5× sequence coverage (Kirkness et al. 2003), is not yet available in searchable form. Fish are at the root of vertebrate diversification, and extant fish are separated from mammals by 450 Myr of independent evolution. The compact vertebrate genome of Fugu rubripes was proposed as a tool for discovery of vertebrate genes and their regulatory elements (Brenner et al. 1993), and, indeed, its draft genome sequence showed significant conservation of protein-coding genes, intron-exon structures, and gene arrangement with human (Aparicio et al. 2002). Partial genomic sequence information is also available for green pufferfish Tetraodon nigroviridis (http://www.genoscope.cns.fr/), which diverged from Fugu 20 to 30 MYA, and also for zebrafish Danio rerio (http://www.ensembl.org/), which diverged 100 to 150 MYA from pufferfish and is a popular experimental model in developmental studies. The availability of genome information has led to increased recent interest in zebrafish as a model system for human disease (Ward and Lieschke 2002).

Here, we compare genomic sequences, including adjacent genes, protein alignments, and phylogenetic analysis of mammalian PRNP, related fish genes, and SPRN genes from mammals and fish, together with their adjacent genes. We provide detailed descriptions of the mammalian PRNP, PRND, PRNT, and SPRN genes, and performed “phylogenetic footprinting” on the aligned human, mouse, and Fugu SPRN genes to define possible regulatory regions and transcription factor–binding sites. We present evidence for novel fish genes related to PrP that we identified in public databases: stPrP-2 from Tetraodon and stPrP-3 from zebrafish. We derived the SPRN sequence coding for Sho experimentally from Tetraodon, and for Fugu, Tetraodon, carp, and zebrafish, we found in silico a duplicated SPRN gene (SPRNB) encoding the related Shado2 protein (Sho2). We have assessed orthology of the mammalian PRNP and fish genes related to PrP and the mammalian SPRN and homologous fish genes, respectively, and have inferred evolutionary patterns for the mammalian PRNP and SPRN genes. We have developed an evolutionary model consistent with our various findings, as well as information available in the literature, and propose it as a framework for investigating gaps in knowledge in these genes, particularly in fish and tetrapod lineages.

Methods and Materials

Public Databases, Searches, and Search Tools

Detection of PRNP Genes and Genomic Environment in Human, Mouse, and Rat

The PRNP genes in human (chr20p13: 4614996–4630236 bp), mouse (Chr2F3: 132911892–132940089 bp), and rat (Chr3Q36: 112889678–112890442 bp) were found by key word search of the Ensembl human (version 12.31.1), Ensembl mouse (version 12.3.1), and Ensembl rat (version 11.2.1) genome databases, respectively. The local genomic environment is also evident from the interactive web genome browser, as is annotation of the genomic sequence.

Detection of Fugu stPrP-1–Coding Gene and stPrP-2–Coding Genes and Their Genomic Context

We identified stPrP-1–coding gene and stPrP-2–coding gene sequences and their local genomic context in the Fugu interactive Ensembl genome browser (version 12.2.1) by using the sequences of their open reading frame (ORFs) as search query (GenBank accession numbers AY141106 and AY188583, respectively) and the server’s BlastN (Altschul et al. 1990) search tool. The local genomic environment and annotation of the genomic sequence is also evident from this Web service. The genes encoding stPrP-1 and stPrP-2 are located on scaffold_96 and scaffold_155, respectively.

Identification of Tetraodon Sequences Containing PrP-like and stPrP-2–Coding Genes

We identified genomic contig FS_CONTIG_4238_2 containing the Tetraodon PrP-like and stPrP-2–coding genes by using the sequence of the Tetraodon PrP-like ORF (Suzuki et al. 2002) as search query and the BlastN program provided in the Genoscope Web service. We identified overlapping genomic clone FS_CONTIG_4238_1 by using the terminal 200 bp of FS_CONTIG_4238_2 as search query. In addition, two more overlapping clones (FS_CONTIG_42895_1 and FS_CONTIG_31286_1) were identified by using the same strategy. Sequences of these contigs were merged into a virtual contig (Tetraodon virtual contig 1) of length 22,249 bp. We verified this assembly using the PiPMaker program (Schwartz et al. 2000) and alignment to the orthologous Fugu genomic sequence. The sequence was also annotated by using the NIX interactive Web tool (Williams, Woollard, and Hingamp 1998; http://www.hgmp.mrc.ac.uk/NIX). The TestPrP-2 accession number is BN000527.

Identification of Zebrafish Coding Sequence for stPrP-1 (C-Terminal Only) and stPrP-3

To identify the zebrafish stPrP-1–coding gene, we used nucleotide sequence from the Fugu stPrP-1 ORF as search query and the BlastN Web service from the Ensembl zebrafish interactive genome database (version 14.2.1). We detected coding sequence for the C-terminal part of the protein only on chromosome fragment ctg10456. A later search (version 22.3b.1) detected the complete gene and genomic context on chr10 (11,350,000 to 11,390,000 bp). Genomic sequence containing the zebrafish stPrP-3 gene (assembly_234, NA3274.1) was identified by using the Fugu stPrP-2-coding sequence as search query and the BlastN Web service from the Ensembl zebrasfish genome DB, as above. The local genomic environment and its annotation are evident from the interactive genome browser. The ZePrP-3 accession number is BN000526.

Detection of SPRN Genes and Genomic Environment in Human, Mouse, Rat, Fugu, and Zebrafish

We described recently identification of the SPRN gene in human, mouse, rat, Fugu, and zebraship (Premzl et al. 2003). We used here an identical approach to extract the SPRN gene sequences and their local genomic environment from the Ensembl databases. The SPRN accession numbers are BN000518 (human), BN000519 (mouse), BN000520 (rat), and BN000521 (Fugu).

Identification of Tetraodon Genomic Sequence Containing the SPRN Gene and Cloning of Sho Coding Sequence

Genomic contig FS_CONTIG_4144_1 was identified in the Genoscope database as described in Premzl et al. (2003). By using its terminal 200-bp sequence as search query and the BlastN program on the Web service, we identified overlapping clone FS_CONTIG_31029_1. Using the same strategy, the next overlapping clone FS_CONTIG_37429_1 was identified. These sequences were assembled into a virtual contig of 19,029 bp (Tetraodon virtual contig 2; we used a 10-kb sequence containing the SPRN gene from this virtual contig in further analyses). We verified this assembly by using the PiPMaker program to align it to the orthologous Fugu genomic sequence and annotated it by using the NIX interactive Web tool. As the low quality of the FS_CONTIG_4144_1 sequence does not allow translation of the Sho ORF, we used its sequence to design primers (5’-CTAAGACATCCCGCATGACAG-3’ and 5’-AGCATCTGCTGCAATCACAC-3’) flanking the Sho ORF, using the MacVector version 7.0 program (Oxford Molecular Group 2000). DNA was extracted from Tetraodon cells using standard procedures (Sambrook, Fritsch, and Maniatis 1989) and used as the template for PCR amplification using FastStart Taq polymerase (Roche) according to the PCR procedure using GC-RICH solution. The PCR product was cloned using the TOPO cloning kit (Invitrogen). The cloned fragment was sequenced by using universal primers M13 forward and M13 reverse and the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems). Products of cycle sequencing reactions were run on an ABI3730 DNA sequencer (Applied Biosystems). The TesPRN accession number is AJ717305.

Identification of Sho2 Coding Sequences in Zebrafish, Fugu, Tetraodon, and Carp

Zebrafish genomic sequence containing the Sho2 (ctg10456) coding sequence was identified using the zebrafish Sho ORF sequence as search query and the BlastN program available from the Ensembl (version 14.2.1) interactive Web service. Genomic sequence containing the Fugu gene encoding Sho2 (scaffold_96) was
identified using the zebrafish Sho2 coding sequence as search query and the BlastN program available from the Ensembl (version 12.2.1) Web genome browser. The Tetraodon genomic contig (FS_CONTIG_41464_1) containing Sho2 coding sequence was identified using the Fugu Sho2 ORF sequence and the BlastN search tool from the Genoscope database. The express sequence tag (EST) (CA964511) containing the ORF for the carp Sho2 was detected by BlastN search of the NCBI est_owners database, using the zebrafish Sho2 coding sequence as search query. Local genomic environment and its annotation are provided in the Ensembl genome browser for Fugu SPRNB and zebrafish sprnb. The SPRNB accession numbers are BN000522 (Fugu), BN000523 (zebrafish), BN000524 (carp), and BN000525 (Tetraodon).

Interspersed Elements Analysis

We determined the relative content of repetitive elements (interspersed repeats, small RNA, satellites, simple repeats, and low-complexity sequence) in the human PRNP, PRND, PRNT, and SPRN genes and in mouse Prnp, Prnd, and Sprn genes by using the RepeatMasker program on its Web server (http://ftp.genome.Washington.edu/RM/RepeatMasker.html) with the high sensitivity, slow option. The genomic coordinates and database source for the genes are given in Supplementary Material Table S1 and in the Ensembl databases listed above. Content and distribution of interspersed repeats in the human, mouse, and rat genomic sequences was also determined using RepeatMasker. Interspersed repeats only were masked in the genomic sequences we used for the VISTA alignments.

VISTA Global Alignments

We used the VISTA Web server (Mayor et al. 2000; http://www-gsd.lbl.gov/VISTA/) to align genomic sequences in two in silico experiments. In the first experiment, we aligned human (chr20: 4,558,866 to 4,938,939 bp), mouse (chr2: 132,862,228 to 133,103,751 bp), rat (chr3: 112,821,949 to 113,040,637 bp), Fugu (chr_scaffold_155: 247,572 to 271,811 bp), and Tetraodon (Tetraodon virtual contig 1 [see above]) genomic sequences containing either the PRNP or the stPrP-2 gene, its whole upstream intergenic sequence, and adjacent genes (gene for PrP-like protein in fishes, PRND in mammals, PRNT in human, and RASSF2 and SLC23A1 in both mammals and fishes). In the second experiment, we aligned human (chr10: 134,322,741 to 134,374,165 bp), mouse (chr7: 130,335,510 to 130,371,625 bp), rat (chr1: 199,669,758 to 199,704,156 bp), zebrafish (assembly_203: 4,494,411 to 4,519,077 bp), Fugu (chr_scaffold_28: 384,496 to 394,338 bp), and Tetraodon (Tetraodon virtual contig 2 [see above]) genomic sequences containing the SPRN gene, its whole upstream intergenic region, and adjacent genes (gene encoding GTP-binding protein in all species, amine oxidase (AO)–coding gene in mammals and pufferfish, and long-chain fatty-acyl elongase–coding gene in zebrafish). Before submission to VISTA, transposable elements in the mammalian sequences were masked by using RepeatMasker as above. Human sequence and its annotation were always used as the base sequence. Pairwise sequence comparisons were calculated with a threshold of 50% identity in a 50-bp window. The minimum identity shown in the VISTA plots is 30%.

PiPMaker Local Alignments

Pairwise alignments of the Fugu and Tetraodon genomic sequences containing the stPrP–2–coding gene and SPRNA (see above), and human (chr20: 4,657,104 to 4,708,670 bp) and mouse (chr2: 132,957,930 to 132,998,136 bp) genomic sequences bordered by PRND and RASSF2 genes, were performed by using default settings of the PiPMaker program available as a Web service (Schwartz et al. 2000; http://bio.cse.psu.edu/cgi-bin/pipmaker?basic). Results are presented as dot plots.

Phylogenetic Footprinting and Promoter Analysis

Phylogenetic footprinting of the human, mouse, and zebrafish genomic sequences containing the SPRN/Sprn gene was performed by using the FootPrinter program as a Web service (Blanchette and Tompa 2002; http://abstract.cs.washington.edu/~blanchem/FootPrinterWeb/FootPrinterInput.pl). We used sequence upstream to the Sho ORF and the whole intergenic region in this experiment. The orthologous rat, Tetraodon, and zebrafish sequence data were not used in this analysis because of gaps in the sequences (rat and zebrafish) or low quality (Tetraodon). We used a conservative approach in this analysis, accepting only motifs detected in all three species and setting low parsimony scores: score 0 for 6-bp and 7-bp motifs, score 1 for 8-bp and 9-bp motifs, and score 2 for 10-bp, 11-bp, and 12-bp motifs. A 13-bp motif (score 2) was identified after inspection of the 12-bp motif sequences. Potential transcription factor–binding sites in these same human and mouse genomic sequences were analyzed by using the MatInspector program available as a Web service (Quandt et al. 1995; http://www.genomatix.de/cgi-bin/eldorado/main.pl). Again, we used a conservative approach in this analysis: motifs were identified with core similarity 1 and matrix similarity score above the optimized matrix similarity score. Motifs predicting transcription factor–binding sites deposited in the TRANSFAC database (Wingender et al. 1996) were identified.

Alignment and Phylogenetic Analysis of PrP-Related and Sho Proteins

Amino acid sequences of 11 PrP and PrP-related proteins (details in figure S3 of Supplementary Material online) and 10 Sho/Sho2 proteins (details in figure S4 of Supplementary Material online) were obtained from GenBank, or as described in this study. The sequences of these two protein sets were aligned independently using ClustalW (Thompson, Higgins, and Gibson 1994), and the result was subsequently refined using genetic data environment (GDE) (Smith et al. 1994). From the PrP and Sho alignments, 118 and 149 sites, respectively, were chosen for phylogenetic analyses using the maximum-likelihood program package MOLPHY (Adachi and Hasegawa 1996). The two alignments were aligned using ProtML (Adachi and Hasegawa 1996) with the JTT-F model of amino acid
substitution. Initially, every possible tree was assessed to produce a set of 2,000 near-optimal trees (using the -jfen 2000 option). This set of trees was then analyzed more rigorously without, and with, nearest-neighbor interchanges (using the -jfu and -jfru options). Local bootstrap probabilities (LBP) were estimated using the resampling estimated log-likelihood method for internal edges in the tree examined. Trees were compared statistically using the Kishino-Hasegawa test (Kishino and Hasegawa 1989), and topological model uncertainty (Wolf et al. 2000) was considered using model averaging (Jerumiin et al. 1997); when considered, we used the class V weighting scheme with $\alpha = 0.05$ to produce (1) a majority-rule consensus tree and (2) relative-likelihood scores (RLS) for internal edges in the inferred phylogeny.

Analysis Software

The Lasergene (DNASTAR, Madison, Wis.) and Vector NTI (InforMax, Frederick, Md.) software packages were used for basic handling and analyses of the nucleotide sequence and protein data. Signal peptide cleavage sites were predicted by using the SignalP program (Nielsen et al. 1997; http://www.cbs.dtu.dk/services/SignalP/) and GPI-anchor sites were predicted by using the bigPI-predictor program (Eisenhaber, Bork, and Eisenhaber 1999; http://mendel.imp.univie.ac.at/gpi/gpi_server.html) available as Web services. Other figures were drawn using Microsoft PowerPoint and Adobe Illustrator.

Results and Discussion

Genomic Environments of PRNP, SPRN, and Related Genes

Sequence comparison is a multistep process (Frazer et al. 2003). We found and extracted genomic sequences from the public databases. We then annotated the reference sequence, aligned the sequences, and identified conserved segments.

Identification of Genomic Sequences Containing PRNP in Mammals and Related Genes in Fish

Using the Ensembl interactive genome browsers for human, mouse, rat, Fugu, and zebrafish (http://www.ensembl.org/) we found the genes and extracted their local contexts. Genomic sequence of Tetraodon containing stPrP-2 and PrP-like proteins was derived by detecting overlapping genomic clones in the Genoscope database (http://www.genoscope.cns.fr/) and assembling them into a virtual genomic contig (Tetraodon virtual contig 1 [see below]). In figure 2, we present gene order and relative orientation in the local genomic regions of mammalian PRNP genes and fish genes encoding stPrP-1, stPrP-2, stPrP-3, and PrP-like proteins. In mammals, genes adjacent to PRNP and PRND are RASSF2 and SLC23A1. The PRNT gene is present in humans but not in rodents (fig. 2A [see also below]). In pufferfish, genes encoding stPrP-2 and PrP-like proteins are also adjacent to Rassf2-coding and Slc23a1-coding genes (fig. 2C). The stPrP-2–coding gene is in the same position and relative orientation with respect to the adjacent genes as is mammalian PRNP, suggesting an evolutionary relationship.

In contrast, the stPrP-1–coding genes from Fugu and zebrafish are located in different genomic environments (fig. 2D and E). In the Fugu genome, stPrP-1 is flanked distally by TA-PPC2 (T-cell activation protein phosphatase 2C) and EPI-64 (epi64 protein). Proximally, it is arranged head-to-head with a gene that appears to be a homolog, SPRNB, of the gene, SPRN, which we recently discovered encoding the protein Shadoo (Sho) (Premzl et al. 2003); SPRNB, in turn, is flanked by KCNIP3 (calsenilin) (see also below). In the zebrafish genome, we detected stPrP-1 adjacent to sprnb, also head-to-head, but flanked, interestingly, by the rassf2 gene proximally and the slk gene (STE20-like kinase) distally (i.e., a different genomic environment from that in Fugu). The proximity in fish of the two genes, stPrP-1 and SPRNB/sprnb, is very interesting, suggesting an evolutionary and possible functional link between stPrP-1 and Sho2 proteins. We have not found the SPRNB gene in mammalian sequence databases. There are several possible explanations; for example, the gene has been lost from the mammalian lineage, or it is so diverged that searches cannot recognize it. The most obvious possibility, that the gene was duplicated after divergence of mammals and fish, is not consistent with phylogenetic results we present later.

We found another stPrP-like gene, we here call stPrP-3, in zebrafish in the Ensembl genomic contig NA3274.1. Its proximal flanking gene is unknown, but its distal flanking genes code for green cone photoreceptor (gcp) and annexin a6 (anxa6).

We observe some major differences between the PrP-related genes in the mammalian and fish genomes. First, neither PRND nor PRNT genes were detected in fish. Although the similarity in position and orientation of PRNT in human and the PrP-like gene in pufferfish might suggest orthology, sequence comparison shows no similarity (i.e., they are different genes). Obversely, no PrP-like protein–coding genes were found in mammals. Suzuki et al. (2002) correctly noted that the genes encoding Rassf2 and Slc23a1 are in proximity to the PrP-like protein–coding gene in Fugu but reported them in a different gene order and relative gene orientation. Rivera-Milla, Stuermer, and Malaga-Trillo (2003) stated that the genomic environments around the Fugu genes encoding PrP461 and PrP-like proteins seemed to have undergone multiple chromosomal rearrangements, but they did not present supporting data. In the Fugu genomic data, Oidtmann et al. (2003) found the stPrP-1–coding gene on scaffold_96, and the stPrP-2–coding gene on scaffold_155, adjacent to the PrP-like protein–coding gene; they concluded it was unlikely that ancestors of the stPrP-1 and stPrP-2 genes evolved directly into present mammalian forms of PRNP and PRND. As supported by our later phylogenetic analysis, we suggest that stPrP-2 shares an ancestor with the fish gene encoding stPrP-1, but the gene duplication giving rise to stPrP-1 and stPrP-2 occurred after the evolutionary separation of fish and mammals. Although stPrP-2 in pufferfish and PRNP in mammals share the same proximity to the Rassf2-coding and Slc23a1-coding genes and are likely orthologous, their
sequences have diverged greatly, and they may have evolved significantly different functions. The relaxation of conserved contiguity between the sites with different genes distal to PRNP and stPrP-2 (i.e. PRND and PrP-like genes) suggests independent duplications producing PRNP-PRND and stPrP-2 and PrP-like genes. As shown in figure 1, the C-terminal regions of PrP and stPrP-2 proteins have similar features, but these are not present in the PrP-like protein sequences; hence, the function of the PrP-like protein is likely different from that of stPrP-2.

Identification of Genomic Sequences Containing SPRN in Mammals and Fish

We recently characterized the genomic environment of the SPRN gene in the human, mouse, rat, Fugu, and zebrafish genomes (Premzl et al. 2003); this is summarized in figure 3, together with that of Tetraodon. Interestingly, we have recently detected an apparent duplication of the SPRN gene in human, about 140 kb downstream, from genomic data (L. Sangiorgio, B. Strumbo, and T. Simonic, unpublished results). We obtained the Tetraodon SPRNA context by finding overlapping genomic sequence reads in the Genoscope database (http://www.genoscope.cns.fr/) and assembling them into a virtual genomic contig (Tetraodon virtual contig 2 [see below]). (The fish gene is here labeled SPRNA to differentiate it from the related gene in fish, SPRNB, reported in this paper.) Genes adjacent to SPRN in mammals and SPRNA in pufferfish are those encoding the GTP-binding protein (GTP) and AO; both gene order and relative orientation are conserved in the two vertebrate clades, implying orthology. However, in zebrafish, although we found the GTP-binding–protein gene adjacent to sprna in the same tail-to-tail orientation, the proximal adjacent gene is the long-chain fatty-acyl elongase (fae)–coding gene rather than the AO-homologous gene. Because the next proximal genes are not known from the assembly, it is not clear whether the AO-encoding gene is present.

Identification and Annotation of Tetraodon Genomic Context Containing stPrP-2–coding Gene and SPRN

We assembled and annotated Tetraodon genomic sequences in silico. First, we extracted 22,249 bp of genomic sequence containing the stPrP-2–coding gene and its neighbor genes (FS_CONTIG_4238_2, FS_CONTIG_4238_1, FS_CONTIG_24895_1, FS_CONTIG_31286_1; Genoscope) and assembled it into the Tetraodon virtual contig 1. To verify this assembly and assess its validity for
our comparative genomic analysis, we aligned it to its orthologous Fugu genomic sequence using the PiPMaker program (Schwartz et al. 2000), which is able to compare both complete and incomplete sequences. The dot plot of the Fugu (chr_scaffold_155: 247,572 to 271,811 bp; Ensembl) and Tetraodon genomic sequences is given in figure S1A in Supplementary Material online.

Among the four genes in this genomic fragment, the exon-intron structure is known from comparison of the genomic and cDNA sequences only for the PrP-like protein–coding gene. The single-exon ORF of the stPrP-2–coding gene and GenScan predictions for Rassf2-coding gene and Slc23a1-coding gene exons are also shown. The dot plot indicates that the Tetraodon virtual contig 1 is assembled in an order consistent with the Fugu genomic sequence; we conclude the assembly is correct and valid for comparative analysis. Additionally, we deduced the Tetraodon stPrP-2 amino acid sequence from the genomic sequence information (see later figure S3 of Supplementary Material online).

Second, we identified 19,029 bp of Tetraodon genomic sequence (FS_CONTIG_4144_1, FS_CONTIG_31029_1, FS_CONTIG_37429_1; Genoscope) containing the SPRN gene and adjacent genes, and merged it into the Tetraodon virtual contig 2. The dot plot of 10 kb of this sequence aligned to orthologous Fugu genomic sequence (chr_scaffold_28: 384,496 to 394,338 bp; Ensembl) is shown in figure S1B of Supplementary Material online; this shows the GenScan exon predictions for the single-exon Sho ORF and for genes encoding AO and GTP-binding protein. The dot plot again shows consistent patterns in the aligned sequences, suggesting correct assembly of the second virtual Tetraodon contig and its suitability for computational cross-species comparisons. Conservation of sequence proximal (~4.5 kb) and distal (~6 kb) to the GTP-binding protein–coding gene may denote exons not recognized by the GenScan prediction (Guigo et al. 2003) (see also figure S2 of Supplementary Material online).

Comparative Genomic Analysis of PRNP and SPRN Genomic Regions

Genomic sequence annotation comprises gene order and relative transcriptional orientation, gene structure, distribution of repeat elements, and distribution of GC islands. The information for the human (base [reference] sequences in our study) and mouse genomes is provided in the interactive Ensembl genome browsers (http://www.ensembl.org/). An extensive collection of transcripts is available only for these two species, with the rat, Fugu, and zebrafish genome annotations (http://www.ensembl.org/) being much less comprehensive. There are some major differences between the genomes of homeothermic and poikilothermic species; for example, the GC compositional heterogeneity in poikilothermic animals is less pronounced (Aparicio et al. 2002). In addition, the depth of fish transposable element analysis is less than that of primates and rodents (Aparicio et al. 2002). In the following section, we compare annotations of the human and mouse PRNP,

![Fig. 3.—Overview of the local genomic contexts of the SPRN gene coding Shadoo protein in mammals and fish. AO, Amine oxidase; GTP, GTP-binding protein; FAE, long-chain fatty-acyl elongase. Figure is approximately to scale as shown by rulers. For ruler under (D), gene sizes and intergenic distances refer to Fugu. Genomic coordinates and DB information for the genomic sequences roughly correspond to those used for the cross-species Vista analysis; these are given in Materials and Methods.](https://academic.oup.com/mbe/article-abstract/21/12/2210/1071063)
PRNP, PRND, PRNT, and SPRN genes and of their local genomic contexts.

Gene Structure, Gene Features, Gene Density, and CpG Islands

Genomic location, gene structure, gene size, GC content, and features of exons and introns of human and mouse PRNP, PRND, PRNT, and SPRN genes are summarized in table S1 of Supplementary Material online.

We found that gene density and GC content are much higher in the SPRN genomic environment than in the PRNP environment, in all mammals analyzed (figs. 2 and 3, and see also tables S2 and S3 of Supplementary Material online). There are five genes in 380,074 bp of the human PRNP local genomic environment, which is 45.02% GC rich, compared with three genes in 51,425 bp of the human SPRN gene context, which is 50.66% GC rich. In all four genes, (PRNP, PRND, PRNT, and SPRN), the ORF is contained in a single coding exon. Their gene lengths correlate inversely with GC content, and GC content is also higher in the exons (comparing exon 1/intron 1 pairs and, where relevant, exon 2/intron 2 pairs). The GC content of the human and mouse SPRN genes (66% and 58%, respectively) is much higher than that of PRNP (42% and 45%), PRND (46% and 47%), and PRNT (43%, human). For reference, mammalian GC content is known to vary genome-wide at different scales; for example, the average GC content of the human genome is 41%, ranging from 36% to 47% on the large scale (<10 Mb) and from 33% to 59% on the small scale (<300 kb) (Lander et al. 2001), with similar values reported for the mouse genome (Waterston et al. 2002). An analysis of CpG islands in the human and mouse genomic contexts of PRNP, PRND, PRNT, and SPRN using the cpgplot program (Larsen et al. 1992) (results not shown) showed islands associated with the PRNP and SPRN promoter regions but not those of PRND (as shown by Cominici et al. [2001]) or PRNT.

Distribution of Transposable Elements in PRNP and SPRN Genes

The role of repeats in the genome is controversial. Although often considered as just “junk,” it is now recognized that active elements have reshaped genomes (e.g., creating new, modified, or reshuffled genes). Passive elements are excellent markers for mutation and selection analyses (Lander et al. 2001). Differences in the repeat content and distribution denote differences in the gene/genome dynamics (Thomas et al. 2003). Discernable interspersed repeats are reported to comprise approximately 45% of the human genome (Lander et al. 2001) and 37.5% of the mouse genome (Waterston et al. 2002). The lower content in mouse is thought to be caused by a higher nucleotide substitution rate in mouse (4.5 $\times$ 10$^{-9}$) than in human (2.2 $\times$ 10$^{-9}$), which makes older elements difficult to identify. The depth of the repeat analysis using the RepeatMasker program (http://ftp.genome.Washington.edu/RM/RepeatMasker.html) is 150 to 200 and 100 to 120 MYA for human and mouse elements, respectively (Waterston et al. 2002).

Our analysis of the distribution of interspersed repeat elements in the local genomic environments of the human, mouse, and rat PRNP and SPRN genes shows more transposable repeats in human than in mouse and rat, for both regions (given in tables S2 and S3 of Supplementary Material online). There are more interspersed elements in the SPRN genomic environment than in that of PRNP of human and rat but fewer in mouse.

At the single-gene level, however, we observe some striking differences in the transposable-element content and distribution; this is summarized in table 2 for the human and mouse PRNP, PRND, PRNT, and SPRN genes. For a detailed analysis of repeat distribution in eutherian PRNP and PRND genes, see Lee et al. (1999) and Cominici et al. (2001). The repeat content of these four genes correlates with their length and inversely with their GC content; thus, PRNP, has the most repeats. Lee et al. (1999) suggested that the gene had expanded independently in all lineages since the mammalian radiation by numerous insertions. Strikingly, in contrast, SPRN, the shortest gene with the highest GC content, is devoid of transposable elements.

The frequency of fixation of transposable elements is known to vary genome-wide. At two extremes, human Hox gene clusters contain less than 2% repeats in approximately 100 kb, whereas human Xp11 has accumulated 89% of transposable elements in a 525-kb region (Lander et al. 2001). There is also a strong correlation between divergence in noncoding DNA and the amount of repetitive DNA (Chiaromonte et al. 2001): “flexible” genomic regions accumulate many changes, whereas “rigid” regions accumulate fewer. Rigidity of sequence may reflect strong selection on a large number of gene

<p>| Table 2 |
| Summary of Transposable-Element Content in Human and Mouse PRNP, PRND, PRNT, and SPRN Genes |</p>
<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>Length (bp)</th>
<th>SINE %</th>
<th>LINE %</th>
<th>LTR %</th>
<th>DNA %</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>PRNP</td>
<td>15,241</td>
<td>4.56</td>
<td>40.66</td>
<td>0.00</td>
<td>0.88</td>
<td>46.10</td>
</tr>
<tr>
<td></td>
<td>PRND</td>
<td>6,549</td>
<td>18.48</td>
<td>5.63</td>
<td>0.00</td>
<td>0.00</td>
<td>24.11</td>
</tr>
<tr>
<td></td>
<td>PRNT</td>
<td>9,387</td>
<td>21.39</td>
<td>4.08</td>
<td>3.84</td>
<td>4.80</td>
<td>34.11</td>
</tr>
<tr>
<td></td>
<td>SPRN</td>
<td>3,912</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mouse</td>
<td>PRNP</td>
<td>28,198</td>
<td>5.98</td>
<td>3.62</td>
<td>25.35</td>
<td>0.36</td>
<td>35.31</td>
</tr>
<tr>
<td></td>
<td>PRND</td>
<td>5,269</td>
<td>5.11</td>
<td>0.00</td>
<td>0.00</td>
<td>3.89</td>
<td>9.00</td>
</tr>
<tr>
<td></td>
<td>PRNT</td>
<td>2,203</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

NOTE.—Gene sequence coordinates are given in table S1 of Supplementary Material online.
regulatory elements (Lander et al. 2001), or, alternatively, may be determined by the local genomic mutation rate (Chiaromonte et al. 2001). We suggest that the fact that SPRN lacks repeats in its local genomic environment is an indication of strong purifying functionally driven selection acting on the gene. Conversely, PRNP’s flexibility and “promiscuity” for accepting repeat insertions suggests a more relaxed evolutionary history of the gene.

Alignment of Genomic Sequences

Orthologous genes can usually be aligned and recognized by comparative genomic analysis in closely related species. Between evolutionarily more distant species, such as mammals and fish, it is the coding regions that are primarily recognizable (Frazer et al. 2003). However, where rapid divergence of nucleotide sequence, indels, and gene loss or acquisition has occurred, sequences cannot readily be aligned, so orthologs will not be recognized (Kellis et al. 2003). Indeed, the analysis of Thomas et al. (2003) showed that almost one third of human coding sequence did not align to fish in the corresponding genomic region.

To detect highly diverged but orthologous sequences in two long contiguous sequences, global alignments are an advantage (Frazer et al. 2003). We used the VISTA global alignment tool (Mayor et al. 2000) for this purpose to detect conserved, orthologous elements in mammalian and fish genomic sequences. Cross-species VISTA analysis of genomic fragments containing the mammalian PRNP and fish stPrP-2–coding and PrP-like protein–coding genes is shown in figure 4, with similar analysis for the mammalian and fish SPRN genes in figure S2 of Supplementary Material online. Description of the
genomic context and coordinates is given in Materials and Methods.

VISTA Plots for Genomic Regions of PRNP, PRND, and PRNT

We observe substantial conservation in both coding and noncoding sequences of the PRNP, PRND, RASSF2, and SLC23A1 genes between human and rodents; conservation is high in the exons, and there are a few highly conserved regions in the introns as well. For the PRNP gene, our results agree with those of Lee et al. (1998). For the PRND gene, we note highest conservation in the coding exon 2: exons 1 and 3 seem less conserved. An intergenic exon in mouse (85,944 to 86,046 bp in the mouse sequence) does not align with its corresponding human region. The human and rodent RASSF2 and SLC23A1 genes align extensively.

We observe that conservation with rodents in the human PRNT gene region differs from that of other genes shown in the plot. There is almost no conservation between rat and human (none at all in exons), and conservation between human and mouse appears poor.

In contrast to the human-rodent comparison, the VISTA plot in figure 4 detects no conservation between mammalian PRNP and fish stPrP-2–coding and PrP-like protein–coding genes. Neither homology criteria nor non-homology criteria for gene orthology are fulfilled. First, the coding exons of the genes do not align, indicating divergence of mammalian PRNP and fish stPrP-2–coding genes beyond detectable conservation. Second, there is clear evidence of deletion, translocation, or duplication events in the local genomic sequence since divergence of mammals and fish: the PRND gene (and PRNT gene in human) exists in mammals but not in fish, whereas the PrP-like protein–coding gene is present in fish only. The adjacent RASSF2 and SLC23A1 genes align with their fish orthologs in the exons.

PipMaker Analysis of PRNT

To test the Vista observation for the PRNT gene further, we aligned human (chr20: 4,657,104 to 4,708,670 bp;
Ensembl) and mouse (chr2: 132,957,930 to 132,998,136 bp; Ensembl) genomic sequence regions between the PRND and RASSF2 genes using the PipMaker program. The dot plot results presented in figure 5, like those of the Vista results, show no conservation of PRNT gene exons between human and mouse.

When we translated exon 2 of the (human) PRNT gene, we detected two potential ORFs encoding 60 and 94 residues and also several smaller ORFs (not shown). Makrinou, Collinge, and Antoniou (2002) reported the expression of the PRNT gene exclusively in testis, as for the PRND gene, and noted 50% similarity and 42% identity between a potential 94-residue protein and human Dpl. The absence of a long ORF and its apparently recent origin suggest that human PRNT is a pseudogene arising originally from duplication of PRND. Pseudogenes are often remnants of duplicate genes arising either from tandem duplication or retrotransposition of processed mRNA. Although their sequences show similarities to coding regions of known proteins, they have acquired many stop codons or frame shifts so that they no longer code for full-length protein. They are usually not transcribed but occasionally are resurrected in response to a new organisal need (Harrison and Gerstein 2002), or they acquire additional functions such as a specific regulatory role (Hirotsume et al. 2003). Also, in the intron of the PRNT gene, we detected a processed pseudogene sequence with high homology to the mRNA for isopentenyl-diphosphate delta isomerase 1, which is disrupted by an Alu insertion.

It is likely that the PRNT gene appeared in the human lineage after the evolutionary split with rodents. This is consistent with the position of the PRNP genomic context on chromosome 20 in a highly recombinogenic region that may have fostered a recent duplication of PRND to create the human-specific PRNT gene. However, it is also possible that the PRND duplication is more ancient and that PRNT survives as a pseudogene in other mammalian lineages as well as human but has been deleted in rodents.

VISTA Plot for SPRN Genomic Region

Alignment of the mammalian and fish genomic regions containing the SPRN and adjacent genes shows
Phylogenetic Footprinting of the SPRN Gene

“Phylogenetic footprinting” is a method for identifying regulatory elements in a single gene by comparing orthologous sequences from several species (Blanchette and Tompa 2002). The basic assumption is that functional regions, including regulatory elements, are under greater selective pressure and will be more conserved between species than nonfunctional regions. However, as regulatory elements may be quite short (5 to 20 bp), it is difficult to recognize them in the nonfunctional background noise. We used the program Footprinter (Blanchette and Tompa 2002; http://abstract.cs.washington.edu/~blanchem/FootPrinterWeb/FootPrinterInput.pl) to identify such elements conserved in the human, mouse, and Fugu SPRN genes. The program reports sets of conserved motifs, taking into account a phylogenetic tree relating the input species.

As shown in table 3, we identified 16 conserved motifs upstream to the SPRN ORF, in the intron, exon 1, and upstream promoter. In human and mouse, five motifs were detected in the upstream promoter, one in exon 1, and 10 in the intron, as shown in figure 6. Some motifs are duplicated. Although this set of motifs contains candidates for regulatory regions of the SPRN gene, it may also contain false positives. Also, Footprinter may miss motifs present in a single species (i.e., false negatives), motifs shorter than 6 bp in multiple species, motifs containing indels, motifs that fail to meet statistical significance, and dimers with variable internal sequences (Blanchette and Tompa 2002). We next checked whether any known transcription factor–binding sites were among the detected motifs using the MatInspector program (Quandt et al. 1995; http://www.genomatix.de) for human and mouse sequences. This program predicts transcription factor–binding sites deposited in the TRANSFAC database (Wingender et al. 1996). We detected 155 and 159 likely transcription factor–binding sites in the human and mouse sequences, respectively (69 or 82 in the intron, 11 or 14 in exon 1, and 75 or 62 in the upstream promoter): full lists for human and mouse are given in tables S4 and S5, respectively, of the Supplementary Material online.

This combined analysis showed that three of the motifs detected by phylogenetic footprinting correspond to predicted transcription factor–binding sites, and, significantly, we note that these elements are present in the same relative order in the human and mouse sequences. All three motifs were detected in the SPRN intron; we note, again, that this intron is free of transposable elements, so the presence of the motifs suggests the intron may play an important role in gene regulation.

Motif 4 denotes the binding site for the nurrl nuclear receptor, which is expressed only in brain and is known to play an important role in coordinate neuroendocrine regulation of activity of the hypothalamic/pituitary/adrenal axis (Murphy and Conneely 1997). It is also critical for dopaminergic neuron development by activating tyrosine hydroxylase transcription in a cell context–dependent manner (Kim et al. 2003). Abrerrations in the dopaminergic system are associated with Parkinson disease and schizophrenia.

Motif 14 corresponds to the activating transcription factor 6 (ATF6) binding site. ATF6 is a member of the

Conservation of the SPRN-coding exon between mammals and fish satisfies homology criteria for gene orthology. Next, conserved gene order and relative transcription orientation of SPRN and its adjacent gene encoding GTP-binding protein (including also AO-coding gene in pufferfish) between mammals and fish indicates that no rearrangement occurred in this genomic fragment after the evolutionary divergence of fish and mammals 450 MYA. Thus, we conclude that the SPRN gene is likely to be orthologous between mammals and fish.
basic leucine-zipper family. It is involved in induction of the endoplasmic reticulum (ER) stress response, during which transcription of genes encoding molecular chaperones and folding enzymes located in the ER is upregulated. Some genes in this pathway are directly activated by ATF6 binding sites. Upstream to ATF6 in the ER stress response is IRE1 (Wang et al. 2000). Remarkably, the ER stress-response pathway is involved in familial Alzheimer disease (FAD) pathogenesis. FAD-linked PS1 mutants attenuate autophosphorylation of IRE1 and lead to disease (FAD) pathogenesis.

### Table 3

Conserved Motifs in Human, Mouse, and Fugu SPRN Gene Identified by Phylogenetic Footprinting

<table>
<thead>
<tr>
<th>Motif (Length in bp) and Localization in Human/Mouse SPRN</th>
<th>Position in Human/Mouse/Fugu SPRN Genomic Sequence Relative to Sho ORF</th>
<th>Score $^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TGGAGA (6) P/P</td>
<td>−1290−1285/−1301−1296/−1143−1138</td>
<td>0</td>
</tr>
<tr>
<td>2. GCTGAG (6) E1/P</td>
<td>−890−885/−1049−1044/−886−881</td>
<td>0</td>
</tr>
<tr>
<td>3. GCTGAA (6) I/I</td>
<td>−658−653/−574−569/−776−771</td>
<td>0</td>
</tr>
<tr>
<td>4. TCCAGA (6) I/I</td>
<td>−466−461/−434−429/−303−297</td>
<td>0</td>
</tr>
<tr>
<td>5. GAACCC (6) I/I</td>
<td>/−58−−53/−276−271/−63−57</td>
<td>0</td>
</tr>
<tr>
<td>6. ATCTCC (6) I/I</td>
<td>−36−31/−292−287/−100−95</td>
<td>0</td>
</tr>
<tr>
<td>7. CTTTCC (6) I/I</td>
<td>−1411−1404/−1479−1472/−1453−1446</td>
<td>1</td>
</tr>
<tr>
<td>8. GAGGAGCA (8) P/P</td>
<td>−1346−1339/−1262−1255/−1500−1493</td>
<td>1</td>
</tr>
<tr>
<td>9. TGAACACAA (8) P/P</td>
<td>−933−926/−1011−1004/−927−920</td>
<td>1</td>
</tr>
<tr>
<td>10. GGAGGGCT (8) P/EI</td>
<td>−1027−1022/−/−</td>
<td>1</td>
</tr>
<tr>
<td>11. GGAGGTG (8) I/I</td>
<td>−782−775/−826−819/−780−773</td>
<td>1</td>
</tr>
<tr>
<td>12. CCAGCCA (8) I/I</td>
<td>−668−661/−786−779/−759−752</td>
<td>1</td>
</tr>
<tr>
<td>13. CCAGGCTAa (9) I/I</td>
<td>−202−194/−388−380/−209−201</td>
<td>1</td>
</tr>
<tr>
<td>14. GcGTGAC (10) I/I</td>
<td>−304−297/−</td>
<td>1</td>
</tr>
<tr>
<td>15. TG GCCGT (10) P/P</td>
<td>−161−154/−/−</td>
<td>1</td>
</tr>
<tr>
<td>16. CCCuCAAGCCCT (13) I/I</td>
<td>−361−352/−423−414/−291−282</td>
<td>2</td>
</tr>
<tr>
<td>17. CCCuCAAGGCT (10)</td>
<td>−/−1390−1404/−1411−1421/−1453−1446</td>
<td>2</td>
</tr>
<tr>
<td>18. CCCuCAAGGC (10)</td>
<td>−/−1404−1479−1472/−1453−1446</td>
<td>2</td>
</tr>
</tbody>
</table>

Note.—Localization corresponds to position in human and mouse SPRN intron (I), exon 1 (E1), and upstream promoter (P).

| Motif | Capital letters indicate conserved position in motif; lowercase letters indicate variable position in motif. |
| Motif | Motif duplications were identified manually after inspection of results. |
| TRANSFAC matrix sequences identifying potential transcription factor–binding sites were included in MatInspector program analysis: motif was required to be detected both in human and mouse sequences. |
| Parsimony score corresponds to main motif. |
| Motif denotes VS NBRE.01 matrix sequence for Nurr1-binding motif. |
| Motif denotes VS ATF6.02 matrix sequence for activating transcription factor 6–binding motif. |

### Figure 6

Potential regulatory motifs in human SPRN and mouse Sprn genes identified by phylogenetic footprinting. Motifs 4, 14, and 16, labeled by an asterisk (*), denote potential nurr1, ATF6 (activating transcription factor 6) and MAZR (MYC-associated zinc-finger protein–related) transcription factor–binding sites, respectively.
impaired induction of the ER stress response. These mutants also attenuate the ATF6 signaling pathway (Kudo et al. 2002). We have already shown predominant expression of mammalian SPRN in brain (Premzl et al. 2003). These new findings of conserved and clustered nur1 and ATF6 transcription factor–binding site motifs in the human and mouse introns are further suggestive of Sho’s brain-specific function and are targets for experimental analysis.

Finally, part of the third conserved motif 16 binds the MYC-associated zinc-finger protein–related transcription factor (MAZR). MAZR interacts with Bach2, a B-cell and neuron-specific transcription repressor (Kobayashi et al. 2000).

Protein Alignment and Phylogenetic Analysis

As already discussed for the genomic sequences and summarized in figures 1–3 and table 1, we have expanded the data set of fish proteins related to PrP. First, we deduced a sequence of 395 amino acids for Tetraodon stPrP-2 protein from the assembled genomic sequence. Second, we translated the zebrafish stPrP-3–coding sequence in Ensembl genomic contig NA3274.1 into a protein of 561 amino acids. Third, based on the incomplete genomic sequence we had assembled, we cloned and sequenced the Tetraodon Sho ORF and deduced a 155-residue protein, thus adding a third fish Sho sequence to those for zebrafish and Fugu we had previously reported (Premzl et al. 2003). Last, as already discussed, we identified a new Sho-related class, Shadoo2 (Sho2), of sequences in public databases. We were able to deduce Sho2 protein sequences of 150, 150, and 135 amino acids, respectively, from genomic information for Fugu (scaffold_96 from Ensembl), Tetraodon (FS_CONTIG_41464_1 from Genoscope), and zebrafish (ctg10456 from Ensembl). Carp Sho2 was conceptually translated from EST data (CA964511 from NCBI) to give a protein of 145 amino acids.

As described in Materials and Methods, the amino acid sequences of 11 PrP-related proteins from human, chicken, turtle, Xenopus, Fugu, Tetraodon, and zebrafish and 10 Sho and Sho2 proteins from human, mouse, rat, Fugu, Tetraodon, zebrafish and carp, were aligned independently using ClustalW (Thompson, Higgins, and Gibson 1994). After refinement of the alignments using GDE (Smith et al. 1994), we assessed the alignments and chose 118 and 149 sites from the PrP-related and Sho/Sho2 proteins, respectively, for phylogenetic analyses using the maximum-likelihood program MOLPHY (Adachi and Hasegawa 1996). The alignments are shown in figures S3 and S4 of Supplementary Material online.

The analysis of the PrP-related sequence set identified a single most-likely tree, shown in figure 7, and 281 near-optimal trees, none of which differed significantly from the most-likely tree. The total tree length is 5.65, implying that every site in the alignment (figure S3 of Supplementary Material online) has changed on the average 5.65 times, which, in turn, implies that interpretation of the tree must be done with some caution; this is also reflected in several low LBP and RLS values (fig. 7). However, the fact that we found only 282 “good” trees (i.e., the most-likely tree and near-optimal trees) out of 34,459,425 possible trees places the result in a better light.

The most-likely tree groups the human, chicken, turtle and frog sequences together to the exclusion of all fish sequences, as expected from known phylogeny. Although the inferred evolutionary relationship among higher vertebrates has human PrP more distantly related to birds and reptiles than to frog, the most likely tree is not
significantly different from others consistent with the current view on tetrapod evolution (i.e., with PrPHu and PrPFr changed over). The divergence of the amino acid sequence between the fish proteins related to PrP and the PrPs of higher vertebrates suggests no orthology between these proteins. The clustering pattern of the stPrPs also indicates that the genes coding for these proteins were duplicated in fish after the evolutionary split from tetrapods.

The analysis of the Shadoo proteins identified a single most-likely tree, shown in figure 8, and 48 near-optimal trees, none of which differed significantly from the most likely tree. The total tree length is 4.77, implying that every site in the alignment (figure S4 of Supplementary Material online) has changed on the average 4.77 times. Again, this implies caution must be applied in interpretation of the tree, this need being reflected also in some low LBP and RLS values (fig. 8). However, again we note we found only 49 “good” trees (i.e., the most-likely tree and the near-optimal trees) out of 2,027,025 possible trees, setting the result in a better light.

The Shadoo tree indicates that Shos and Shos2s lie on two separate branches, implying these two genes duplicated before the divergence of fish from tetrapods. Most importantly, the fish Shos cluster with their mammalian homologs, rather than with their fish Sho2 paralogs. This clustering pattern strongly suggests orthology between mammalian and fish Shadoos, an inference previously made on the basis of sequence conservation (Premzl et al. 2003). As noted previously, conserved genomic context and failure to detect SPRNB in mammalian genomes suggests either that the SPRNB gene has been lost between fish and mammals or that it is highly diverged and yet to be found.

It is established that gene duplication occurred in several fish lineages and that many duplicated fish genes have only one homolog in mammals (Taylor et al. 2003; Aparicio et al. 2002). Two fates of duplicated genes have been proposed. The classical model of neofunctionalization predicts that one of the duplicate loci retains its original function, whereas the other duplicate is fixed only if rare beneficial mutations occur (Ohno 1970). This model fits current knowledge of Shos and Sho2s, with the Sho2 duplicate either deleted in mammals (and other tetrapods?) or so highly diverged by neofunctionalization that its mammalian ortholog is not recognizable. The alternative model proposes that both duplicates are preserved because of subfunctionalization, where proteins encoded by the duplicates complement each other functionally (Force et al. 1999). This model fits the fish stPrPs, which have sequences similar to each other but dissimilar to those of tetrapod PrPs.

Protein Sequence Features

We introduced previously a model for the structures of mature Sho, PrP, and PrP-related proteins, which is a useful basis for discussing both conserved and highly variable regions in these sequences (Premzl et al. 2003). The model consists of four regions: a basic region 1, which shows a tendency for insertion of repeat or other sequence (region 2), a hydrophobic region 3, and a C-terminal
region 4. Figure 1 and figures S3 and S4 of Supplementary Material online show the model represents a quite dynamic structural scaffold, with substantial differences in the presence, lengths, and sequence compositions of the regions and insertions among the complete Sho/PrP set, particularly taking account of the numerous fish proteins. As we have already compared the sequence features of all proteins in figure 1 except for the ones newly reported here (TestPrP-2, ZestPrP-3, TeSho, FuSho2, TeSho2, CaSho2, and ZeSho2 [see table 1]), we confine attention here largely to the new proteins.

Fish proteins from the known stPrP set (FustPrP-2, 424 residues; TestPrP-2, 395 residues; FustPrP-1, 461 residues; and ZestPrP-3, 561 residues) are much longer than tetrapod PrPs (frog, 216 residues; turtle, 270 residues; chicken, 273 residues; and human, 253 residues) or fish PrP-like proteins (~170 to 190 residues) and show significant sequence heterogeneity among themselves, especially in the seemingly "free format" low-complexity repeats and large insertion in region 2 and for large insertions within region 4. Most of the difference in length between TestPrP-2 and FustPrP-2 is in the large insertion in region 2, whereas the very large ZestPrP-3 protein shows increases of approximately 90 and approximately 45 residues in the region 2 insertion and in region 4, respectively, compared with stPrP-2s. More useful alignment and analysis requires better knowledge of the number of stPrP proteins in individual fish and their variation among the major fish lineages (see also Protein Alignment and Phylogenetic Analysis above). As shown in figure 1, the new stPrPs are also predicted to have N-terminal and C-terminal signal sequences for extracellular export and GPI-anchor attachment, respectively (Premzl et al. 2003 and Materials and Methods), and also one disulfide bridge and three N-glycosylation sites.

The main significantly conserved feature among all PrP-related proteins is the hydrophobic region 3 (residues 112 to 131 in HuPrP) and also subsequent sequence in region 4 up to about residue 160 (HuPrP). Interestingly, this includes the beginning of sequence shown to be folded in the NMR structures (L125-R228 for HuPrP [Zahn et al. 2000]) and includes the first unusual (highly hydrophilic) α-helix (D144-M154, HuPrP) and the first part of the antiparallel β-sheet (Y128-G131, HuPrP). This segment also includes the region from residues 119 to 138 (HuPrP) implicated as a PrP$_{106-126}$ dimerization site and potential binding site for PrP$_{106-126}$-PrP$_{106-126}$ complex formation (Horiiuchi and Caughey 1999; Zuegg and Greedy 2000). We note a repetitive insertion in this region in fish PrP-like proteins.

Compared with PrP-related proteins, conservation of features, including length, is much higher in the Sho/Sho2 set. We previously noted very high conservation (identity 81% to 96%) among mammalian Shos (human, mouse, and rat), and good conservation, particularly for zebrafish, between fish and mammalian sequences to slightly beyond the end of the hydrophobic sequence (identity 41% to 53%; zebrafish 1 to 78) (Premzl et al. 2003). The fish proteins are all of similar length (FuSho, 146 residues; TeSho, 149 residues; ZeSho, 131 residues; FuSho2, 150 residues; TeSho2, 150 residues; ZeSho2, 136 residues; CaSho2, 145 residues). There is an insertion in the Fugu and Tetraodon Sho basic repeats that is not present in other Shos. Of note, however, is that Sho2 sequence in this region is different from that in Shos, with the Sho basic region N-terminal to the hydrophobic region being missing in Sho2s. Although there are some local regions of sequence conservation, including around the N-glycosylation site conserved in all Shos and Sho2s, the C-terminal regions of fish Shos and Sho2s are quite diverged; this is also the sequence region most diverged between fish and tetrapod Shos (see above).

Model for the Evolution of PRNP-Related and SPRN-Related Genes

In their paper reporting the finding of the fish PrP-like protein–coding genes, Suzuki et al. (2002) made the initial suggestion of an evolutionary link between the fish PrP-like protein–coding genes and tetrapod PRNP genes. This was based on both weak "homology" (both encode extracellular GPI-anchored proteins with repeats and an unusual internal hydrophobic region [see figure 1]) and context-based (conservation of genes encoding Rasf2 and Sle23a1 in both genomic contexts) criteria. The report by Oidtmann et al. (2003) of the stPrP-2–coding gene proximal to the PrP-like protein–coding gene in Fugu, somewhat strengthened the homology argument for an evolutionary relationship between tetrapod PRNPs and this fish gene because the C-terminal regions of the proteins are similar, whereas those of PrP-like proteins are quite different (see figure 1). However, this finding weakened the context-based argument. Although the orientation of the stPrP-2 gene is now the same as that of mammalian PRNP, the two genes are separated from the Rasf2 coding gene by different genes (PrP-like gene and PRND) with different orientations (fig. 2). We have suggested previously that this could be accounted for by independent duplications. Oidtmann et al. (2003) also reported another fish gene, stPrP-1, in another genomic context. In our earlier paper (Premzl et al. 2003) reporting the finding of the SPRN gene in both fish and mammals, we provided clear evidence, both homology and context based, that suggested a direct evolutionary relationship. However, at this stage, we had no evidence for an evolutionary link between Sho and PrP proteins, although the overall structural features were intriguingly similar.

In this paper, we applied several methods of homology-based (VISTA genomic global alignment, promoter-region footprinting, and protein sequence alignment and phylogenetics) and context-based (genomic context and relative gene order and orientation) analysis to see whether more definitive statements can be made on the relationships between the fish and tetrapod genes. We also included in the analyses sequences of several relevant new genes we found in fish. It is now clear that fish contain a plethora of genes with similarities to PRNP and SPRN, with quite diverged sequences even among the few fish lineages available to us. Also, data on SPRN and PRND genes of tetrapods other than mammals are lacking. Remarkably, in all these genes, the ORF is contained within a single exon. Hence, although we can make some firm conclusions, others are more tentative and await the correction of the current draft data for Fugu, Tetraodon,
and zebrafish genomes as well as the availability of more data for tetrapod lineages other than mammals (i.e., amphibians, reptiles, birds, marsupials, and monotremes).

Considering first the issues for PrP and related genes, our VISTA analysis for global alignment of genomic environment does not show conservation between mammalian PRNP and fish stPrP-2 and PrP-like protein–coding genes, although the adjacent Rassf2 and Slc23a1 coding genes align with their fish orthologs in the exons. There is clear evidence of rearrangements in this genomic region, as the PRND gene (and PRNT gene in human) exists in mammals but not in fish, where the PrP-like protein–coding gene is so far reported only in fish. It, thus, seems clear that the PRNP and stPrP-2–coding genes have diverged greatly since they last shared a common ancestor. There is no basis for suggesting functional homology; that is, the genes are likely to have evolved significantly different functions. Second, our phylogenetic analysis suggests that the gene duplication giving rise to stPrP-1/stPrP-3 and stPrP-2 is likely to have occurred after evolutionary separation of fish and mammals. The clustering of the fish stPrPs separately from the tetrapod PrPs in figure 7 is indicative of the divergence of the amino acid sequence between the fish and higher-vertebrate proteins, and of likely different functions.

Remarkably, we found a new SPRN-related gene, SPRNB, adjacent to the stPrP-1 coding gene in both Fugu and zebrafish. As current data indicate the genomic context of the genes (fig. 2) are different, caution in interpreting these results is necessary. However, it seems unlikely to be accidental that these two genes have been found together twice. Hence, these findings provide the first hint of an evolutionary link between the SPRN and PRNP families. The fact that these two genes code for the unique hydrophobic segment, as we reported initially (Premzl et al. 2003), is further evidence at the protein level of an evolutionary link.

Considering now the SPRN/SPRNβ genes, a quite different picture emerges. All evidence suggests mammalian and fish SPRN/sprn genes are likely to be orthologous; that is, directly descended from a common ancestor and with preserved function. VISTA analysis shows the SPRN gene aligns in our cross-species analysis, together with its adjacent genes, whereas protein-sequence alignment and phylogenetic-tree analysis also indicate conservation between mammalian and fish Shos. Phylogenetic analysis shows Shos and Sho2s lying on two separate branches (fig. 8), implying the two genes were duplicated before the divergence of fish from tetrapods.

We present an evolutionary model consistent with these various findings in figure 9, as a basis for discussion and future investigation. The model proposes that the ancestral gene leading to all the PrP-related and Sho-related encoding genes was the SPRN-like gene. First, an ancient prevertebrate duplication produced the SPRNA and SPRNB genes within an environment that may have contained the AO-binding protein and GTP-binding protein encoding genes proximally, and the Rass2 and Slc23a1 encoding genes distally. The model then proposes separation of the SPRN and SPRNB genes, by a translocation of half of the gene cluster. The subsequent history of the two branches suggests the genomic environment containing the SPRNB gene was highly recombinogenic, whereas that containing the SPRNA gene was stable, leading to the currently known fish and mammal Sho orthologs. The stability of this gene suggests that orthologous SPRN genes coding Sho protein will be found in the other tetrapod lineages. A further duplication of the SPRNB gene is then proposed, still before the divergence of fish from tetrapods, with the duplicate genes acquiring additional C-terminal coding sequence to produce SPRNB1 and SPRNB2 proteoglycans with a completed C-terminal domain, as in tetrapod PrP and fish stPrP-1/-2/-3 proteins (fig. 1).

As shown in figure 9, acquisition of the additional sequence to form the complete C-terminal domain at this stage is necessary to explain subsequent gene evolution steps. It is intriguing that this model predicts the C-terminal domain sequence of genes evolved from SPRNB1 and SPRNB2 has been truncated or replaced at least three times, leading to the PrP-like protein and Sho2 (SPRNβ) coding genes in fish and the PRNT gene in human. Figure 9 shows that SPRNB2 translocated at this ancestor stage as separate contexts is consistent with the known fish stPrP-1 and stPrP-2 gene contexts (fig. 2). However, as no genes descend from SPRNB2 are reported in tetrapods, it is not certain if this translocation occurred before or after the separation of fish from tetrapods.

After divergence of fish from tetrapods, the model proposes independent duplications of the SPRNB1 and SPRNB2 proteoglycans. In the tetrapod lineage, a gene duplication of SPRNB1 produced a gene cluster containing PRNP and PRND genes, together with Rassf2 and Slc23a1 coding genes. The timing of this duplication is unclear because known genes are limited to PRNPs in frogs, reptiles, birds, and mammals and PRND in mammals. As noted above and elsewhere in this paper, it is not known at what stage the SPRNB2 gene was deleted in tetrapod evolution or whether it has simply diverged beyond levels currently detectable in mammals. The model allows for a more ancient origin of PRNT than necessary simply to explain the existing minimal data (i.e., absent in rodents but present in human).

In the fish branch, initial duplications of the SPRNB1 proteoglycans to produce stPrP-2 and PrP-like protein–coding genes and duplications of the SPRNB2 proteoglycan to produce SPRNB (Sho2 coding gene, as observed) and stPrP-1–coding gene, are proposed. As drawn in figure 9, these gene clusters are already separated. If the separation occurred after these duplications, translocation of the SPRNB and stPrP-1 fragment might more conveniently explain the different contexts observed in Fugu and zebrafish. The additional complexity and plethora of genes in fish is consistent with existing knowledge of the higher rate of gene duplication in fish and the fact that many duplicated fish genes have only one homolog in mammals (Taylor et al. 2003; Aparicio et al. 2002).

Dynamics of Mammalian PRNP and SPRN Genes

We observed some interesting differences in the mammalian SPRN and PRNP gene characteristics, which
offer some insights into how these genes have been evolving. First, although in the SPRN and PRNP local genomic environments (covering several genes as in VISTA analyses in figure 4 and in figure S2 of Supplementary Material online), we found the percentage of repetitive elements is comparable (tables S2 and S3 of Supplementary Material online), at the single-gene level the picture is strikingly different. The SPRN gene has no transposable elements, whereas PRNP and PRND have up to 46% and 24%, respectively, in human, all in the introns (table 2). Consistent with this picture of gene dynamics, the brain-specific SPRN gene has evolved conservatively between fish and mammals, and Sho is predicted to have a conserved basic function in vertebrate brain.

By contrast, we have shown that the PRNP gene has undergone major and rapid changes at several levels of comparison: it does not align with sequences of fish stPrP-2–coding genes (fig. 4); the sequence, and subsequence composition (Strumbo et al. 2001), of the tetrapod PrPs (frog, reptile, bird, and mammal) show significant differences; and the number of exons and size of introns varies significantly in mammals (Lee et al. 1998; Makrinou, Collinge, and Antoniou 2002 [table 2]). These features are characteristic of genes evolving under relaxed evolutionary constraints, typical of proteins with rapidly evolving function (Harrison and Gerstein 2002; Kitami and Nadeau 2002). What these functions are has yet to be elucidated. Despite extensive study, the functions of mammalian PrP remain elusive (Aguzzi and Hardt 2003), and no functional studies, or tissue expression profiles, of the other tetrapod PrPs have been reported.

Conclusions

Evolution can be broadly viewed as an interplay between conservation and change that enables perpetuation of life: whereas maintenance of organization requires conservation, variation allows adaptation (Radman, Matic, and Taddei 1999). In this context we suggest that PrP and Sho are two components of a complex brain system, with distinct evolutionary histories and functional roles. Whereas the conserved Sho appears to have a conserved role in vertebrate brain (“maintenance”), the relaxed evolution of PrP may have fine-tuned its adaptive role in vertebrate brain in a lineage-specific fashion. The intriguing similarities in the overall structures for PrP and Sho, taken together with our new model for an ancient common evolutionary origin, reinforce our earlier suggestion (Premzl et al. 2003) that the functions may be partly overlapping. However, the fact that our model highlights other lineage-specific PrP-related genes, such as those encoding PrP-like protein, stPrP-1, and Sho2 in fish or Dpl in mammals, suggests PRNP and stPrP-2 genes may have developed lineage-specific functions through coevolution with these genes. It is possible that PrP’s functions over vertebrate evolution have been coevolving with several varied partners and that it has acquired, and lost, a range of relatively weak or subtle, but useful, functions. In this
sense PrP is “dispensable” (Hirsh and Fraser 2001), which allows its remarkable adaptation.

**Supplementary Material**

Figure S1 + caption.pdf

Figure S1. PipMaker dot plots of the genomic fragments containing the stPrP-2-coding and PrP-like protein-coding, and SPRNA genes in Fugu and Tetraodon.

Figure S2 + caption.pdf

Figure S2. VISTA plot for SPRN showing peaks of similarity between human versus mouse, human versus rat, human versus Fugu, human versus Tetraodon, and human versus zebrafish.

Figure S3_S4 + captions.pdf

Figure S3. Alignment of the 10 Sho/Sho2 proteins used in the study.

Figure S4. Alignment of the 10 Sho/Sho2 proteins used in the study.

SupplTablesS1_S2_S3.pdf

Table S1. PRNP, PRND, PRNT, and SPRN genes in human and mouse.

Table S2 and S3. Repeat elements in human, mouse, and rat PRNP and SPRN genes, respectively.

SupplTablesS4_S5.pdf

Table S4. MatInspector analysis of human SPRN regulatory regions.

Table S5. MatInspector analysis of mouse Sprn regulatory regions.

**Acknowledgments**

J.E.G. and J.A.M.G. are supported by the ANU IAS block grant, and J.A.M.G. also by a grant from the Australian Research Council. T.S. is supported by EC grant QLK5-2002-00866 and FIRST. We thank DNASTAR for an extended trial of their Lasergene software. The authors thank Dr. Frances Shannon, JCSMR, ANU and Professor M. Radman, Faculté de Médecine Necker-Enfants Malades, Université Paris V for helpful discussions. Tetraodon cells were kindly supplied by Dr. Frank Grutzner, Research School of Biological Sciences, ANU. This is research paper #006 from the Sydney University Biological Informatics & Technology Centre.

**Literature Cited**


Williams, G. W., P. M. Woollard, and P. Hingamp. 1998. NIX: a nucleotide identification system at the HGMP-RC. URL: http://www.hgmp.mrc.ac.uk/NIX/


Takashi Gojobori, Associate Editor

Accepted August 3, 2004