Supporting Information

Arnegard et al. 10.1073/pnas.1011803107

SI Text

Specimen Collection. We collected many of the mormyrid electric fish for this study from two major river basins in west-central Africa during July and August 2006: the Congo River basin (Lékoli River in Odzala National Park, Republic of the Congo) and the Ogooué River basin (Ivindo River near Makokou and Okano River near Mitzig, both in Gabon). We captured adult fish alive and recorded their electric organ discharges (EODs) in water taken from the site of capture. Methods for capturing mormyrids and recording their electric signals are described by Sullivan et al. (1), Arnegard et al. (2, 3), and Lavoué et al. (4), who also provide additional information about our collection sites. After recording the EOD, we euthanized each fish with an overdose of the anesthetic MS222 (tricaine methanesulfonate) and immediately extracted and preserved samples of myogenic electric organ (EO_{myo}) and skeletal muscle (SM). By carefully dissecting the caudal peduncle we were able to obtain samples of EO_{myo} lacking SM. Samples of SM lacking EO_{myo} were taken from the dorsal trunk musculature anterior to the pelvic fin. Tissue samples were preserved in RNAlater (Ambion) and stored in a freezer at −20 °C, or on ice, for the duration of all fieldwork in Africa. All but one of the euthanized specimens were tagged with permanent specimen numbers and fixed in 10% formalin (phosphate buffered) as they lay flat in a tray of the fixative for 2 wk. These specimens were then transferred to 70% ethanol for long-term storage and deposited in the Cornell University Museum of Vertebrates. The exception was a large specimen of *Mormyrops anguilloides* (#6200), which was too cumbersome to export from Africa. Table S1 gives collection sites, specimen numbers, and museum catalog numbers for all specimens collected in Africa. Four mormyroid species were obtained through the aquarium trade and processed as described above: *Gymnarchus niloticus* (donated by M. Kawasaki, University of Virginia, Charlottesville, VA); and *Petrocephalus soudanensis*, *Pollimyrus adspersus*, and *Campylomormyrus numenius* (donated by C. D. Hopkins, Cornell University, Ithaca, NY). We also obtained a sample of SM from a close out-group to the Mormyroidea, *Xenomystis nigri*, which we acquired through the aquarium trade. Philip K. Stoddard (Florida International University, Miami) donated specimens of three gymnotiform species, which he purchased from the aquarium trade: *Brachyhypopomus pinnicaudatus*, *Gymnotus cylindricus*, and *Rhamphichthys marmoratus*. We obtained the rest of the gymnotiform species directly from an aquarium trade dealer. We took EO_{myo} lacking SM from *G. niloticus* and all gymnotiforms by sampling this tissue as close as possible to the posterior tip of the caudal fin or filament. Phil Stoddard also provided EOD recordings for all gymnotiform species, which had previously been recorded using the method described by Stoddard et al. (5). Fig. S1 shows EODs of all of the electric fish species sampled for this study.

cDNA Cloning and Sequencing. Upon returning from Africa, field-collected samples were stored at −80 °C and processed for cloning and sequencing within 2 months of being collected. Aquarium trade specimens were euthanized as described above and processed immediately after obtaining them. We isolated RNA from homogenized tissue by guanidinium thiocyanate–phenol–chloroform extraction (6) using RNA STAT-60 (Tel-Test). RNA concentrations were estimated by absorbance readings using the ND-1000 spectrophotometer (NanoDrop Technologies), and quality of the extracts was confirmed by gel electrophoresis. mRNAs coding for $Na_v1.4a$ and $Na_v1.4b$ (transcripts of the paralogs *Scn4aa* and *Scn4ab*, respectively) were reverse-transcribed and amplified by PCR. We cloned the resulting cDNAs into chemically competent *Escherichia coli* by means of the pCR2.1-TOPO vector (6) (Invitrogen). Inserts were sequenced in both directions using the dideoxy chain termination method with an ABI 3100 sequencer (Applied Biosystems). In addition to newly sequenced species, we extended all sequences of Zakon et al. (7) to as many as 2,913 bp (971 codons). All species for which new sequence data were collected for the present study are listed in Table S1, which provides GenBank accession numbers for the analyzed *Scn4aa* and *Scn4ab* transcripts. We supplemented these data with published sequences available for the following species: longtail knifefish (*Sternopygus macrurus*; GenBank accession nos. AF378144 and AF378139); electric eel (*Electrophorus electricus*; GenBank accession no. M22252); threespine stickleback (*Gasterosteus aculeatus*; Ensembl protein coding genes ENSGACT00000004631 and ENSGACT00000019623); zebrafish (*Danio rerio*; GenBank accession nos. DQ149506 and DQ149505); medaka (*Oryzias latipes*; Ensembl protein-coding genes ENSORLT00000014569 and ENSORLT00000024364); spotted green pufferfish (*Tetraodon nigroviridis*; GenBank accession nos. DQ221251 and DQ221252); and panther pufferfish (*Takifugu pardalis*; GenBank accession no. AB030482).

Tissue-Specific Expression Assays. We used a PCR-based approach to compare levels of *Scn4aa* and *Scn4ab* expression in SM and EO_{myo} (when present) in several teleost species. With the exception of the threespine stickleback (see below), mRNAs for these comparisons were extracted from the same individuals used for cloning and sequencing. We stored extracts at −80 °C until we used them for the expression assays. For each species and tissue type assayed, 0.5 μg of total mRNA was reverse-transcribed to cDNA with a specific RT primer. In a subsequent reaction, cDNA templates for *Scn4aa* and *Scn4ab* were amplified by PCR. We designed RT and PCR primers to be paralog- and species-specific in light of our sequence data and available data for other genes in the vertebrate *Scna* gene family (8). We optimized the annealing temperatures and numbers of PCR cycles so that we obtained strong signals without saturation of PCR products. Tables S2 and S3 provide specific RT and PCR primers and PCR cycling conditions for *Scn4aa* and *Scn4ab*, respectively. Amplified fragments were 422–1,163 bp, depending on the gene and species (Tables S2 and S3). In all cases, we ensured that the amplified regions of these paralogs contained at least one intron–exon boundary to exclude amplification of genomic DNA at target fragment sizes. Use of primer sets for both *Scn4aa* and *Scn4ab* simultaneously (in separate reactions) provided us with an internal positive control for mRNA quality and primer utility.

In the many closely related EO_{myo}-possessing species, we found that *Scn4aa* transcripts were always absent (or extremely weakly expressed) in SM; instead, this paralog was essentially exclusively expressed in EO_{myo}. With some genes, alternative splicing (9) could impact the interpretation of gene expression assays such as ours. However, we note that the primers that we developed for each paralog had priming sites that varied substantially across the α-subunit, depending on the species, making alternative splicing an unlikely explanation for this extremely consistent pattern of *Scn4aa* expression. Moreover, splice variants have not been reported for orthologous type IV Na^+ channel α-subunits in other vertebrates (10–12). Therefore, we do not believe that alternative splicing confounded our characterization of expression patterns.

The threespine stickleback used for our expression assay came from Big Mud Lake on Vancouver Island, British Columbia, Canada (provided by D. I. Bolnick, University of Texas, Austin,

TX). Limited cloning and sequencing of *Scn4aa* and *Scn4ab* from this stickleback population were needed to design suitable primers due to sequence differences relative to the draft genome sequence for a *G. aculeatus* individual from Bear Paw Lake, Alaska (13).

Sequence Alignments. We aligned nucleotide sequences separately for *Scn4aa* and *Scn4ab* using the ClustalW routine (14) in BioEdit version 7.0.5.3 (15). We then carefully checked the entire alignments by eye and found them to be completely unambiguous except for a hyper-variable portion of the large intracellular linker between domains II and III: codons 1053–1215 in the electric eel reference sequence (16). Alignment proved questionable in this region, so we excluded this segment of the linker from all analyses. Codons 1052 and 1216 formed the boundaries of the retained sequence because they were the first codons in the questionable region in the N- and C-terminal directions, respectively, to code an invariant amino acid across both the $Na_v1.4a$ and $Na_v1.4b$ alignments (e.g., see Fig. S4, part 2, where a vertical line indicates the excluded sequence). Four alignments were prepared after removal of the hyper-variable sequence: one for the *Scn4aa* gene (30 sequences × 2,469 characters), one for its *Scn4ab* paralog (28 sequences × 2,457 characters), one that was a concatenation of *Scn4aa* and *Scn4ab* (29 sequences × 4,926 characters, hereafter "*Concat*"), and one that consisted of only the third codon positions of *Concat* (29 sequences × 1,642 characters, hereafter "*Concat3rd*"). When concatenating the individual gene sequences, the different taxon sets had to be accommodated in the following ways. The single sequence representing the Ivindo River type I and type II morphs of the "*magnostipes*" complex of *Paramormyrops* were eliminated due to their extremely low divergence from the single sequence representing Okano River type I and type III morphs of this same complex. In addition, the *Scn4ab* gene sequence for *Campylomormyrus numenius* was treated as missing data.

Phylogenetic Inference. We inferred phylogenetic trees from each of the four alignments with the maximum-likelihood (ML) inference software GARLI (17). A beta version that implements partitioned substitution models was used (GARLI-PART version 0.97b, http://www.nescent.org/wg_garli/Partition_testing_version). Bayesian tree inference was also performed with MrBayes version 3.1.2 (18).

Model selection. We selected nucleotide substitution models and partitioning schemes in a multistep procedure using the Akaike Information Criterion (AIC) model selection criterion (19). First, MODELTEST version 3.7 (20) was run on several subsets of the alignments. Six subsets consisted of the individual first-, second-, and third-codon positions of each gene; two consisted of the combined first and second positions of each gene; three consisted of the first, second, and third positions pooled across genes; and two consisted of the entire sequence of each gene. Next, we evaluated three potential partitioning schemes for each gene: all codon positions combined [scheme (123)]; grouped first- and second-codon positions [scheme (12) (3)]; and a scheme allowing independent models for each codon position [scheme (1) (2) (3)]. We evaluated a larger number of partitioning schemes for the concatenated alignment, including partitioning by gene, by codon position, and by both gene and codon position (see Table S4 for a full list). For the *Concat3rd* alignment we evaluated two partitioning schemes, one with independent models for the third positions of each gene and one with a single model applying to both genes (i.e., unpartitioned). In all cases, the model chosen by MODELTEST for a given subset of the alignment was used, with no linking of model parameter values across subsets. Independent rate multipliers were inferred for all partition subsets, allowing different mean substitution rates in addition to rate heterogeneity distributions used within each subset. We then performed full heuristic searches using GARLI-PART for each alignment and partitioning. We ran four replicate searches beginning from rough ML stepwise-addition starting trees. We used the automatic termination condition and default values for other settings. Finally, we compared the best scoring solution of the four searches for each partitioned model using the AIC criterion. The model giving the lowest AIC score was chosen as the best partitioning scheme, and the tree inferred using it was taken as the ML estimate for that alignment.

Clade support values. We performed 1,000 nonparametric bootstrap replicates for each alignment with GARLI-PART, using the best-fitting partitioning scheme. All free model parameters were estimated during each bootstrap replicate. Search conditions were identical to the other GARLI analyses except that only one search replicate was performed per bootstrap resampling. Branches were collapsed in the final trees for each replicate (creating polytomies) if they were inferred to have the minimum allowed length (1×10^{-8} substitutions per site).

We obtained Bayesian support values for the four alignments with MrBayes, using the same partitioning scheme and models chosen in the ML analyses. Models not implemented in MrBayes (e.g., the TVM model) were changed to the more general general time-reversible (GTR) model. We ran four simultaneous independent analyses in MrBayes, each with four differentially heated Markov Chain Monte Carlo (MCMC) chains, sampling every 100 generations. As in the GARLI partitioned analyses, we inferred overall rate multipliers for each partition subset. Other MCMC settings and prior distributions were set at their default values. Preliminary experiments showed that stable likelihood scores were consistently reached by 2×10^4 generations and that the average SD of split frequencies consistently reached 0.01 well before 1×10^5 generations. Final analyses were run to 1×10^6 generations, with 1×10^5 generations discarded as burnin. Post-burnin samples from the four independent runs were pooled for a total of 3.6×10^4 posterior samples per alignment.

Phylogenetic results. The models that we selected using MODELTEST for the alignment subsets appear in Table S5. All were either the GTR model or a slight restriction of that model with a single estimated transition rate (termed TVM). Modeling of rate heterogeneity using both discrete Γ-distributed rate heterogeneity (+G) and a proportion of invariable sites (+I) was chosen for all subsets. We ran the GARLI-PART searches on multiple computers systems; thus, runtimes are not directly comparable. Regardless, the set of four replicate searches performed for each partitioning scheme required less than an hour apiece. For the *Concat, Concat3rd*, and *Scn4aa* alignments, there was no variation in the trees inferred across the four search replicates for each partitioning scheme. In the case of *Scn4ab*, we observed replicate-to-replicate variation (indicating the existence of local topological optima), but we found the same best tree multiple times for all partitioning schemes. The preferred partitioning scheme for both individual genes allowed independent models for each codon position. For the *Concat* alignment, the preferred scheme contained four subsets: first-codon positions of both genes pooled, third-codon positions of both genes pooled, and independent second-codon positions for each gene. The *Concat3rd* alignment preferred an unpartitioned model. AIC scores and rankings of the partitioning schemes are provided in Table S4.

The 1,000 bootstrap replicates were done in parallel across multiple computers and required 43.8, 44.2, 132.4, and 20.7 processor hours total for the *Scn4ab*, *Scn4aa*, *Concat*, and *Concat3rd* alignments, respectively. Bootstrap support values for most branches were quite high for the *Scn4aa*, *Concat*, and *Concat3rd* alignments, but significantly lower for *Scn4ab*. The mean support values across the branches of the trees were 81.6%, 95.1%, 96.7%, and 92.8% for the *Scn4ab*, *Scn4aa*, *Concat*, and *Concat3rd* alignments, respectively. MrBayes analyses required 8.7, 8.8, and 16.1 processor hours for the first three respective alignments. The *Concat3rd* analysis was run on a different computer and required 16.3 processor hours. MCMC convergence in all analyses was unproblematic. For all alignments, the average SD of split frequencies was less than 0.005. Bayesian posterior probabilities were greater than or equal to the corre-

sponding bootstrap values for every branch in every tree and only fell below 0.99 for a handful of branches.

The ML trees inferred for each alignment appear in Fig. S2 and Fig. S3; note that there is very little topological conflict between trees inferred for the independently evolving genes *Scn4aa* and *Scn4ab*. After accounting for the slightly different taxon sets, the ML trees for the *Scn4aa* gene and the *Concat* and *Concat3rd* alignments are all topologically identical. The ML tree for *Scn4ab* is also identical, apart from one weakly supported difference within the clade of Gymnotiformes (South American electric fishes). There, *Gymnotus cylindricus* was placed as sister to the clade consisting of *Rhamphichthys marmoratus*, *Steatogenys elegans*, and *Brachyhypopomus pinnicaudatus*, rather than as a sister to the electric eel (*Electrophorus electricus*), as in the other trees. Despite this single difference, a monophyletic clade of pulse-type gymnotiforms (all of the aforementioned species) was inferred using either the *Scn4aa* or *Scn4ab* datasets (Fig. S3). Relationships among mormyroid taxa (African weakly electric fish species) are identical in the four trees but receive low support from *Scn4ab* and moderately strong support from *Scn4aa*. The Bayesian consensus tree was identical to the ML tree for each alignment.

Disregarding relationships among gymnotiform species, our inferred phylogenies correspond extremely well with the accepted broad relationships among teleost groups (21–29), as well as with relationships within the Mormyroidea and among other Osteoglossomorpha (30–34). In contrast to these congruent relationships, previously there was comparatively little certainty (and molecular evidence) about relationships among major groups within the Gymnotiformes (35–37). To date, our study offers the largest nuclear DNA dataset per gymnotiform species (two different nuclear genes with a total of 4,926 characters) as well as the best-supported phylogenetic hypothesis for the broader, intraordinal relationships among gymnotiforms. In our limited sampling of taxa, this order of electric fishes appears to be composed of the following natural groups. The apteronotids (*Apteronotus albifrons* and *Apteronotus leptorhynchus* in our sample set) form a clade of species with neurogenic adult electric organs (EO_{neuro}); this EO_{neuro}-clade is sister to a clade composed of all other gymnotiform species in our sample set that possess myogenic adult electric organs (EO_{myo}). The basal sister-group relationship of the apteronotids to all other gymnotiforms has been previously proposed on the basis of the pleisiomorphic retention of a caudal fin, present in closely related nonelectrogenic teleosts but absent in all other gymnotiforms (38, 39). The clade of EO_{myo}-possessing gymnotiforms is further split into two natural subclades in our topology, one consisting of only "wave" species and the other containing all "pulse" species included in our study (see Fig. S1 for the distinction between wave and pulse species of weakly electric fish).

Codon Model Analyses. Unless otherwise noted, we performed all codon-based analyses in HyPhy version 1.0 (40). The codon models that we used were generalized versions of the Muse and Gaut model (41). A constrained version of the GTR nucleotide model with three free rate parameters ($r_{AC} = r_{CG}$ and $r_{AT} = r_{GT}$) was used to accommodate nucleotide mutation biases, as chosen by a HyPhy routine that uses the AIC criterion to select among all 203 possible GTR submodels. This model is denoted 012032 in HyPhy parlance. All codon models contained one or more estimated ω-parameters (ω is the nonsynonymous substitution rate divided by the synonymous substitution rate, abbreviated dN/dS in the main text for clarity). In some cases, the models allowed variability in ω across columns of the alignment or branches of the phylogenies. ω-Parameters characterize the selective forces acting on a codon, with small values representing constraint, values near 1 representing selective neutrality, and values greater than 1 representing positive selection. All codon-based analyses presented here require a fixed tree topology as input. In all cases, we used the ML topology inferred using the *Concat* alignment (identical to that inferred using the *Scn4aa* alignment) or subtrees pruned from that tree.

Change in selective forces over time. We expected that the change in expression of *Scn4aa* from SM to solely EO_{myo} (arrows in Fig. 2) would have caused changes in the selective forces acting upon this gene. Such changes in selection pressures might have happened during or shortly after the initial evolution of EO_{myo}, or they might have occurred during the subsequent evolutionary histories of the EO_{myo}-possessing groups. Types of potential change include positive selection, increases in constraint, and decreases in constraint. Each type of change could have affected different codons in the alignment or even the same codon at different times. To investigate these possibilities, we used a number of codon-based methods.

We first looked for evidence of variation in the average selective forces (i.e., ω) acting on the *Scn4aa* and *Scn4ab* genes over the branches of our phylogenies. A codon model assuming a single ω-value applied to all lineages and sites was first fit to the *Scn4aa* and *Scn4ab* trees. Using a likelihood-ratio test (LRT), the fit of this model was compared with the fit of the generally overparameterized "free ratio" model of Yang (42), which infers an independent ω-parameter for each branch of the tree. The test was highly significant for both *Scn4ab* (LRT stat = 208.2; df = 53; $P << 0.0001$) and *Scn4aa* (LRT stat = 819.8; df = 57; $P << 0.0001$), indicating large variations in selective forces over the histories of both genes. Although this unconstrained model is useful as a test for the presence of variability in selection pressure across the branches of a phylogeny, using it to obtain independent ω-estimates for every branch is suspect because of its overparameterization.

GABranch method. To more appropriately investigate when and how selection pressures varied over the evolutionary history of our sampled taxa, we applied the method of Kosakovsky Pond and Frost (43) to the two gene alignments. This method (GABranch) infers three things. The first is a categorization of all branches of a given tree into a number of subsets, each of which share a single ω-value. The second is an estimate of the ω-value of each inferred category. The third is the most appropriate number of such categories. A genetic-algorithm approach is used to search through the space of possible models, with individual models assigned a fitness using the AIC_c criterion. For a given branch, the ω-value of the assigned category applies to all sites. The values assigned to the branches should thus be thought of as average selective tendencies over all sites (codons). Note that if branches are assigned to the same ω-category, this does not necessarily indicate that any particular sites are behaving similarly over the branches in question. Instead, this is simply an indication that the bulk ratio of nonsynonymous-to-synonymous rates is similar among branches assigned to the same ω-category.

This computationally intensive method was carried out using downloadable HyPhy scripts (http://www.hyphy.org/gabranch/) and a parallel MPI (message passing interface) implementation of HyPhy. Five independent replicate analyses were performed for the *Scn4aa* and *Scn4ab* alignments using the Duke University Shared Cluster Resource, with each analysis making use of 16 processors. Specifying this number of processors creates a genetic algorithm population of size 30. After specifying the 012032 nucleotide model, the default GABranch configuration was used, and analyses continued until the automatic stopping criterion was reached. The genetic algorithm samples were postprocessed with the HyPhy script BranchGAResultProcessor.bf. The final results of the independent analyses for each gene were compared to one another for the purpose of verifying that similar solutions had been found. All samples from the five analyses for each gene were then pooled and postprocessed again to obtain the final estimates.

For computational tractability, the GABranch procedure makes use of a very simple codon model that does not allow variation in ω across sites and uses fixed, pre-estimated values of the nucleo-

tide substitution parameters. Thus, the method is useful for inferring partitionings of branches into similar selective categories in a reasonable amount of time, but the estimated parameter values are not ideal. To obtain a better final estimate of the branch lengths and category ω-values, a model was optimized in HyPhy using the branch-partitioning scheme identified by GABranch to model variation in mean ω across lineages in conjunction with discrete Γ-distributed nonsynonymous rate variation across sites (four categories).

GABranch results. The five replicate searches in total required 94.9 h of wall-clock time for *Scn4ab* and 161.9 h of wall-clock time for *Scn4aa*. Factoring in the number of processors used, this amounts to 171.2 processor days for all analyses. The scores of the best solutions found by each of the five runs for each gene were similar, with a range of 4.1 AIC_c for *Scn4ab* and a range of 7.1 AIC_c for *Scn4aa*. For *Scn4ab*, four of five replicates returned a final solution with five ω-categories, and one returned a solution with four categories. All *Scn4aa* replicates returned a final solution with six ω-categories. The samples from the five independent analyses for each gene were pooled and postprocessed to obtain the final results for each paralog.

Reoptimization using the best branch-clustering scheme identified by the GABranch method, along with a model allowing nonsynonymous rate heterogeneity, resulted in an increase in the estimated ω-values for all categories in both genes. The range and distribution of branch ω-values was quite different for the two genes, both before and after reoptimization (Fig. 2). Consider first the *Scn4ab* paralog. As seen in Fig. 2 (*Upper*), the majority of the branches of the *Scn4ab* tree were assigned to low ω-categories. All branches connecting the extant nonelectric lineages fall into one of the two most constrained categories (ω = 0.09 and 0.15). Within the EO_{myo}-possessing clades, most branches show somewhat larger ω-values (ω = 0.22 and 0.44), representing moderate constraint. Two terminal branches for *Scn4ab* possess no inferred synonymous substitutions (*Mormyrops nigricans* and *Paramormyrops* sp. type III). These branches are therefore inferred to have a near infinite dN/dS ratio and are colored gray in the top tree of Fig. 2. Although this result seems artifactual, the existence of this category is justified given the data, and removing it before reoptimization results in worse AIC_c scores.

Next, consider *Scn4aa*, the paralog for which electric fishes exhibit compartmentalized expression in myogenic adult electric organ (EO_{myo}). In this case (i.e., Fig. 2, *Lower*), the partitioning of branches into ω-categories reveals interesting patterns. The branches connecting the nonelectric fishes are assigned to categories with even lower ω-values than in the *Scn4ab* tree (ω = 0.045 and 0.081). Moving up the phylogeny into the EO_{myo}-possessing groups, branches during and/or soon after the change in *Scn4aa* expression are assigned to the highest ω-category inferred for the *Scn4aa* tree (ω = 1.065). Assuming constancy of synonymous rates of substitution over time, this category represents a more than 10-fold increase in mean nonsynonymous rate over that observed in the nonelectrogenic lineages. In the *Scn4aa* tree, the long "hot" branch preceding radiation of the Mormyridae (i.e., all mormyroids except *G. niloticus*) apparently follows the presumed origin of EO_{myo} in this group of electric fishes (arrow) with slight delay. It may be that, along this branch, EO_{myo} originated immediately before the split between the currently monotypic Gymnarchidae and the species-rich Mormyridae or that more amino acid substitutions in *Scn4aa* were required for the origin of a myogenic electric organ specifically adapted for production of pulse-type EODs (the type of signal produced by all Mormyridae) rather than a wave-type EOD (*G. niloticus*; Fig. S1). Regardless, the most parsimonious interpretation of our gene expression data (Fig. 3) is that neofunctionalization of *Scn4aa* occurred nearly concurrently with the origin of EO_{myo} (given the resolution of evolutionary time scales recoverable from extant fishes). This change in expression and the greatly increased rates of nonsynonymous substitution in *Scn4aa* in close temporal proximity to the independent origins of EO_{myo} in both groups of electric fish strongly suggest that *Scn4aa* played a critical role in the origin of this complex novel trait. Understanding the hot terminal apteronotid branch in the *Scn4aa* tree awaits further investigation.

Note that our branch-based analysis is not an explicit test for the presence of positive selection in a lineage because a moderately raised ω-value is consistent with both relaxation of constraint at many sites and positive selection at a smaller number of sites. However, because 395 of 823 columns of the amino acid alignment show no amino acid variability over the entire *Scn4aa* tree (green in the sequence identity plot of Fig. S4), it seems almost certain that at least some sites are under positive selection on these branches.

We present the best-fitting assignments of ω-categories to branches of the *Scn4aa* and *Scn4ab* trees in Fig. 2. However, it should also be noted that other categorizations with similar AIC_c scores were found. Most differed only in the assignment of branches to one of two categories with the lowest ω-values. Fig. 2 should thus be interpreted as a tool for visualizing the location of shifts in selective pressure on the tree. The model-averaged ω-estimates for each branch that are determined by the HyPhy postprocessing script are quite similar to the values assigned by the best-fitting categorization scheme.

Window test methods. For each group possessing adult EO_{myo}, we wanted to identify regions of the sequence that experienced a change in selective forces after the origin of this novel organ of electrical communication. To address this, we developed a parametric window-based method that tests for ω-differences between (*i*) the set of all fish lacking electric organs (the non-EO group—i.e., excluding all fish possessing EO_{myo} or EO_{neuro}) and (*ii*) each group of electric fishes possessing a myogenic adult electric organ (i.e., each of two EO_{myo} groups) in predefined regions of the alignment. The non-EO group was taken as a control (or reference), representing the expected selective pattern while *Scn4aa* continues to be expressed (as $Na_v1.4a$ α-subunits) in skeletal muscle tissue. The method was implemented in custom HyPhy scripts written by D.J.Z. (available upon request).

The full *Scn4aa* alignment was first divided into three subalignments consisting of all of the nonelectric fishes that we sampled (9 spp., N9), all mormyroid electric fishes (12 spp., M12), and all gymnotiform fishes possessing adult EO_{myo} (7 spp., G7). The two *Apteronotus* species were not included in any of the subalignments because their neurogenic adult electric organs made their most appropriate group membership unclear. An unrooted tree topology was obtained for each subalignment by pruning the entire *Scn4aa* tree down to the corresponding set of taxa. These subtree topologies were fixed for all subsequent steps.

The full sequence was next divided into nonoverlapping windows of 10 codons each, with window boundaries shared by all of the subalignments. All windows were contiguous, with no sites skipped between them. Window boundaries were placed such that boundaries of the excluded hyper-variable sequence in the linker between domains II and III (*Sequence Alignments*) fell between two windows rather than (artificially) within a single window. Our determination of window size was a compromise between statistical power and spatial resolution. We achieved too little power with a window size of 5, whereas a window size of 20 offered a spatial resolution across the Na^+ channel genes that we deemed too coarse for optimal interpretation of the results. We excluded sections of sequence less than 10 codons in length at the head and tail end of the alignment. Thus, the first window includes codons 8–17 of our alignment, corresponding to codons 823–832 in the standard electric eel sequence (16).

After defining the subalignments and window boundaries, our method proceeds in four steps, which are performed independently for each EO_{myo}-possessing group. For clarity, we describe a single application of the test to the first window of the G7 subalignment.

Step 1: Global parameter estimation. A simple codon model (as described above) is fit to the full sequence of the N9 and G7 subalignments independently. The parameters estimated by the model are nucleotide substitution parameters, codon frequencies, branch lengths, and a single mean ω-value (i.e., no rate heterogeneity across sites). Some parameter values in the following steps are fixed at the global estimates obtained for each subalignment in this step.

Step 2: Null hypothesis fitting. Alignment columns within the first window are taken from the N9 and G7 subalignments, creating smaller subalignments, termed N9-W_1 and G7-W_1, respectively. A model is simultaneously fit to both N9-W_1 and G7-W_1 using their respective subtrees, estimating a single shared ω-value. This represents the null hypothesis that no change in the selective forces acting on this segment of *Scn4aa* occurred after the evolution of EO_{myo}. No parameters apart from ω are shared between N9-W_1 and G7-W_1. Nucleotide substitution parameters and codon frequencies for each subalignment are fixed at the global values for N9 and G7 inferred in step 1. One additional free parameter is introduced for both N9-W_1 and G7-W_1, a rate-scaling parameter that allows varying total rates of substitution for different windows and subalignments while maintaining the fixed global branch-length estimates for each subalignment estimated in step 1. After finding the parameter values maximizing the likelihood of this model for N9-W_1 and G7-W_1, the sum of the log-likelihood values for the two models represents the score of the null hypothesis, $ln\mathrm{L}(H_0)$.

Step 3: Alternative hypothesis fitting. A model nearly identical to that assumed in step 2 is then fit to N9-W_1 and G7-W_1, allowing each subalignment to assume an independent value of ω. This represents the alternative hypothesis that there was a shift in the selective forces acting on this region of *Scn4aa* after the origin of adult EO_{myo}. No parameters are shared by the subalignments under this model. The sum of the log likelihoods of the two independently optimized models represents the score of the alternative hypothesis, $ln\mathrm{L}(H_A)$.

Step 4: Significance testing. The fit of the two hypotheses may then be compared with a simple likelihood-ratio test. For each window, a *P*-value may be obtained from a χ^2 distribution with a test statistic of $2 \times [ln\mathrm{L}(H_A) - ln\mathrm{L}(H_0)]$ and 1 df. This test procedure was subsequently applied to 81 consecutive window regions along the M12 and G7 subalignments. Note that because N9 is used as a control in the tests of both G7 and M12, the test results are not entirely statistically independent of one another. Because this test is performed independently for each window of 10 codons, a multiple-testing correction must be used. We applied the sequential Bonferroni procedure (44), treating each of the 81 windows of each EO group as an independent test.

In steps 1, 2, and 3, we began parameter optimization from a number of initial values to decrease the chance that parameter estimates became entrapped in local optima. For estimation of the global parameter values from the full sequence in step 1, optimization was begun from five sets of randomly generated parameter values. In steps 2 and 3, three initial values were used for each free parameter: the estimate from the global alignment, $0.25 \times$ the global estimate, and $4.0 \times$ the global estimate. Values for the multiple parameters were used combinatorially, essentially forming a grid centered on the global values. For the H_0 case with three free parameters, this amounted to a $3 \times 3 \times 3$ grid, for a total of 27 sets of starting points. For H_A, the two free parameters for each independently optimized subalignment were chosen from a 3×3 grid, making nine sets.

It should be noted that a significant test result for a given window could be due to relaxation of selection and/or positive selection, either of which likely affect only some fraction of the sites within the window. This method is therefore not a test for positive selection per se, but rather a test for changes in the average selective forces operating within a window. Note also that, although the subalignments share a parameter value under H_0, their corresponding trees are disjoint and no tree branch connects them at any step. Thus, the test deals with how patterns changed after adult EO_{myo} came into existence. The test excludes any changes in selective forces that likely happened on either branch representing the common ancestor of each modern EO_{myo} group, where the shift in expression and the origination of adult EO_{myo} are inferred to have occurred.

Window test results. The window-testing procedure was completed within 2 days using several computers. Multiple parameter optima were observed in a small number of cases, indicating that beginning optimization from multiple values was necessary. The ω-values estimated for each window in each independent subalignment are shown in Fig. 4, where windows with significant *P*-values are indicated with pink circles around the distal tips of the ω-estimate bars. For the gymnotiform subalignment, 13 of 81 window tests were significant, indicating a shift in the average selective forces acting on each of these regions of *Scn4aa* sequence. For the mormyroid subalignment, 19 window tests were significant. Seven window tests were significant for both groups of electric fishes, indicating parallel shifts in selective forces in the same channel regions in these cases.

Visual inspection of Fig. 4 suggests obvious parallelism in terms of which regions of the molecule have significantly changed in the two EO_{myo}-possessing groups relative to the nonelectrogenic reference group. Clearly, it would be of interest to have a formal test of the degree of co-occurrence of significant test windows between the two EO_{myo} groups (i.e., windows showing parallel shifts relative to the selective pattern in nonelectric fishes). Unfortunately, this is made difficult by the fact that the tests of each individual window in each EO_{myo} group are not strictly independent because the test operates by making separate comparisons between each EO_{myo} group and the same non-EO_{myo} reference group. This was necessary because our sample of nine nonelectrogenic teleost species is not large enough to allow accurate estimation of parameters if it were split into separate nonoverlapping subsets to be compared with each electric fish group. Thus, windows with an unusually low estimated ω in the N9 alignment (resulting simply from sampling error in some instances) make a significant window test somewhat more likely for the corresponding window in both G7 and M12. Nonetheless, an approximate test is possible if all windows are assumed to be equally likely to experience a significant change in each group. In this case, a χ^2 test indicates a highly significant co-occurrence ($\chi^2 = 7.97$; df = 1; $P = 0.00477$). This *P*-value will be biased slightly toward increased significance due to the nonindependence of the window tests, but we find it unlikely that the result will become insignificant when an unbiased test of co-occurrence is devised (e.g., when sequences for more nonelectrogenic teleosts become available in the future).

In all windows with significant test results, the change in selective forces was toward higher ω-values, indicating release from constraint or positive selection. Nearly all windows with nonsignificant tests also leaned in the direction of higher ω-values. This suggests that, although some regions are particular hotspots of change under diversifying selection, there has also likely been a decrease in constraint across much of the gene. Global estimates of ω for the three subalignments are consistent with this interpretation. The global ω-values estimated using a single ω-model for the N9, G7, and M12 subalignments are 0.053, 0.212, and 0.293, respectively. Thus, the rate of nonsynonymous substitutions has increased on the order of four or more times relative to the rate of synonymous substitutions in the *Scn4aa* gene after independent origins of the myogenic electric organ (i.e., during the subsequent dramatic diversification of the electrical communication signals in both gymnotiform and mormyroid fishes).

1. Sullivan JP, Lavoué S, Hopkins CD (2002) Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution* 56:597–616.
2. Arnegard ME, Bogdanowicz SM, Hopkins CD (2005) Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. *Evolution* 59:324–343.
3. Arnegard ME, Jackson BS, Hopkins CD (2006) Time-domain signal divergence and discrimination without receptor modification in sympatric morphs of electric fishes. *J Exp Biol* 209:2182–2198.
4. Lavoué S, Arnegard ME, Sullivan JP, Hopkins CD (2008) *Petrocephalus* of Odzala offer insights into evolutionary patterns of signal diversification in the Mormyridae, a family of weakly electrogenic fishes from Africa. *J Physiol Paris* 102:322–339.
5. Stoddard PK, Markham MR, Salazar VL (2003) Serotonin modulates the electric waveform of the gymnotiform electric fish *Brachyhypopomus pinnicaudatus*. *J Exp Biol* 206:1353–1362.
6. Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), 3rd Ed.
7. Zakon HH, Lu Y, Zwickl DJ, Hillis DM (2006) Sodium channel genes and the evolution of diversity in communication signals of electric fishes: Convergent molecular evolution. *Proc Natl Acad Sci USA* 103:3675–3680.
8. Novak AE, et al. (2006) Gene duplications and evolution of vertebrate voltage-gated sodium channels. *J Mol Evol* 63:208–221.
9. Matlin AJ, Clark F, Smith CWJ (2005) Understanding alternative splicing: Towards a cellular code. *Nat Rev Mol Cell Biol* 6:386–398.
10. Plummer NW, Meisler MH (1999) Evolution and diversity of mammalian sodium channel genes. *Genomics* 57:323–331.
11. Raymond CK, et al. (2004) Expression of alternatively spliced sodium channel α-subunit genes. Unique splicing patterns are observed in dorsal root ganglia. *J Biol Chem* 279:46234–46241.
12. Geffeney SL, Fujimoto E, Brodie ED, III, Brodie ED, Jr, Ruben PC (2005) Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. *Nature* 434:759–763.
13. The Broad Institute (2007) gasAcu1.0. Initial assembly of the stickleback genome. Available at http://www.ensembl.org/Gasterosteus_aculeatus/index.html (The Broad Institute, Cambridge, MA). Accessed July 2007.
14. Chenna R, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500.
15. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98.
16. Noda M, et al. (1986) Existence of distinct sodium channel messenger RNAs in rat brain. *Nature* 320:188–192.
17. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation (University of Texas, Austin). Available from http://www.zo.utexas.edu/faculty/antisense/garli/Garli.html.
18. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
19. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
20. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
21. Miya M, et al. (2003) Major patterns of higher teleostean phylogenies: A new perspective based on 100 complete mitochondrial DNA sequences. *Mol Phylogenet Evol* 26:121–138.
22. Saitoh K, Miya M, Inoue JG, Ishiguro NB, Nishida M (2003) Mitochondrial genomics of ostariophysan fishes: Perspectives on phylogeny and biogeography. *J Mol Evol* 56:464–472.
23. Postlethwait J, Amores A, Cresko W, Singer A, Yan Y-L (2004) Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet* 20:481–490.
24. Mank JE, Promislow DEL, Avise JC (2005) Phylogenetic perspectives in the evolution of parental care in ray-finned fishes. *Evolution* 59:1570–1578.
25. Nelson JS (2006) *Fishes of the World* (John Wiley & Sons, Hoboken, NJ), 4th Ed.
26. Steinke D, Salzburger W, Meyer A (2006) Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J Mol Evol* 62:772–784.
27. Lundberg JG, Sullivan JP, Rodiles-Hernández R, Hendrickson DA (2007) Discovery of African roots for the Mesoamerican Chiapas catfish, *Lacantunia enigmatica*, requires an ancient intercontinental passage. *Proc Acad Nat Sci Philadelphia* 156:39–53.
28. Kawahara R, et al. (2008) Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): A new perspective based on whole mitogenome sequences from 75 higher teleosts. *Mol Phylogenet Evol* 46:224–236.
29. Santini F, Harmon LJ, Carnevale G, Alfaro ME (2009) Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol Biol* 9:194.
30. Lavoué S, Bigorne R, Lecointre G, Agnèse JF (2000) Phylogenetic relationships of mormyrid electric fishes (Mormyridae; Teleostei) inferred from cytochrome b sequences. *Mol Phylogenet Evol* 14:1–10.
31. Sullivan JP, Lavoué S, Hopkins CD (2000) Molecular systematics of the African electric fishes (Mormyroidea: teleostei) and a model for the evolution of their electric organs. *J Exp Biol* 203:665–683.
32. Lavoué S, Sullivan JP, Hopkins CD (2003) Phylogenetic utility of the first two introns of the S7 ribosomal protein gene in African electric fishes (Mormyroidea: Teleostei) and congruence with other molecular markers. *Biol J Linn Soc Lond* 78:273–292.
33. Lavoué S, Sullivan JP (2004) Simultaneous analysis of five molecular markers provides a well-supported phylogenetic hypothesis for the living bony-tongue fishes (Osteoglossomorpha: Teleostei). *Mol Phylogenet Evol* 33:171–185.
34. Lavoué S, et al. (2010) Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. *Proc R Soc B*, 10.1098/rspb.2010.1639.
35. Alves-Gomes JA (1999) Systematic biology of gymnotiform and mormyriform electric fishes: Phylogenetic relationships, molecular clocks and rates of evolution in the mitochondrial rRNA genes. *J Exp Biol* 202:1167–1183.
36. Stoddard PK (1999) Predation enhances complexity in the evolution of electric fish signals. *Nature* 400:254–256.
37. Albert JS, Crampton WGR (2005) Diversity and phylogeny of neotropical electric fishes (Gymnotiformes). *Electroreception*, eds Bullock TH, Hopkins CD, Popper AN, Fay RR (Springer, New York), pp 360–409.
38. Triques ML (1993) Phylogeny of the gymnotiform genera (Actinopterygii, Ostariophysi) based on skeletal characters (Translated from Portuguese). *Comunicações do Museu de Ciências e Tecnologia da PUCRS. Série Zoológica* 6:85–130.
39. Gayet M, Meunier FJ, Kirschbaum F (1994) *Ellisella kirschbaumi* Gayet and Meunier, 1991, a gymnotiform fossil of Bolivia and its phylogenetic relationships within the extant Gymnotiformes (Translated from French). *Cybium* 18:273–306.
40. Pond SL, Frost SDW, Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
41. Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11:715–724.
42. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573.
43. Kosakovsky Pond SL, Frost SDW (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol* 22:478–485.
44. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70.
45. Maddison DR, Maddison WP (2000) MacClade 4: Analysis of Phylogeny and Character Evolution. (Sinauer Associates, Sunderland, MA), Version 4.03.

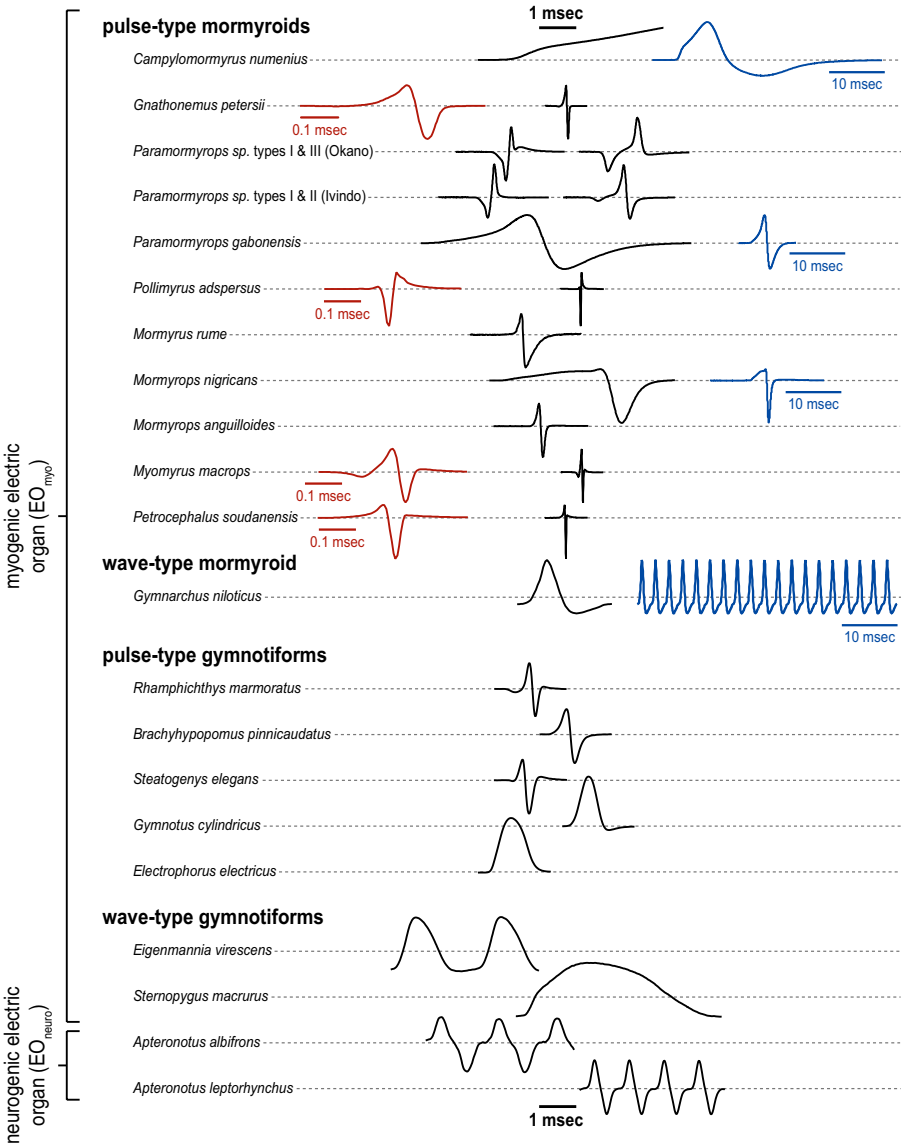

Fig. S1. EOD variation in pulse-type and wave-type mormyroid and gymnotiform fishes. Black traces of EODs are all scaled to the same peak-to-peak amplitude and plotted head-positive-up on the same time scale (scale bars: 1 ms each). Fast or slow EODs are also plotted on alternate (red or blue) scales to better illustrate details of waveform structure (see red and blue scale bars). Most mormyroid EODs were recorded by the authors (see *Specimen Collection*) from the same individuals that were cloned and sequenced for this study. Exceptions include EODs for *Gnathonemus petersii* (provided by G. von der Emde and J. Engelmann), *Mormyrus rume* (P. Moller), and *G. niloticus* (B. A. Carlson). All gymnotiform EODs were kindly provided by P. K. Stoddard, previously recorded as described by Stoddard et al. (5). Table S1 provides specimen information for all EODs recorded by the authors.

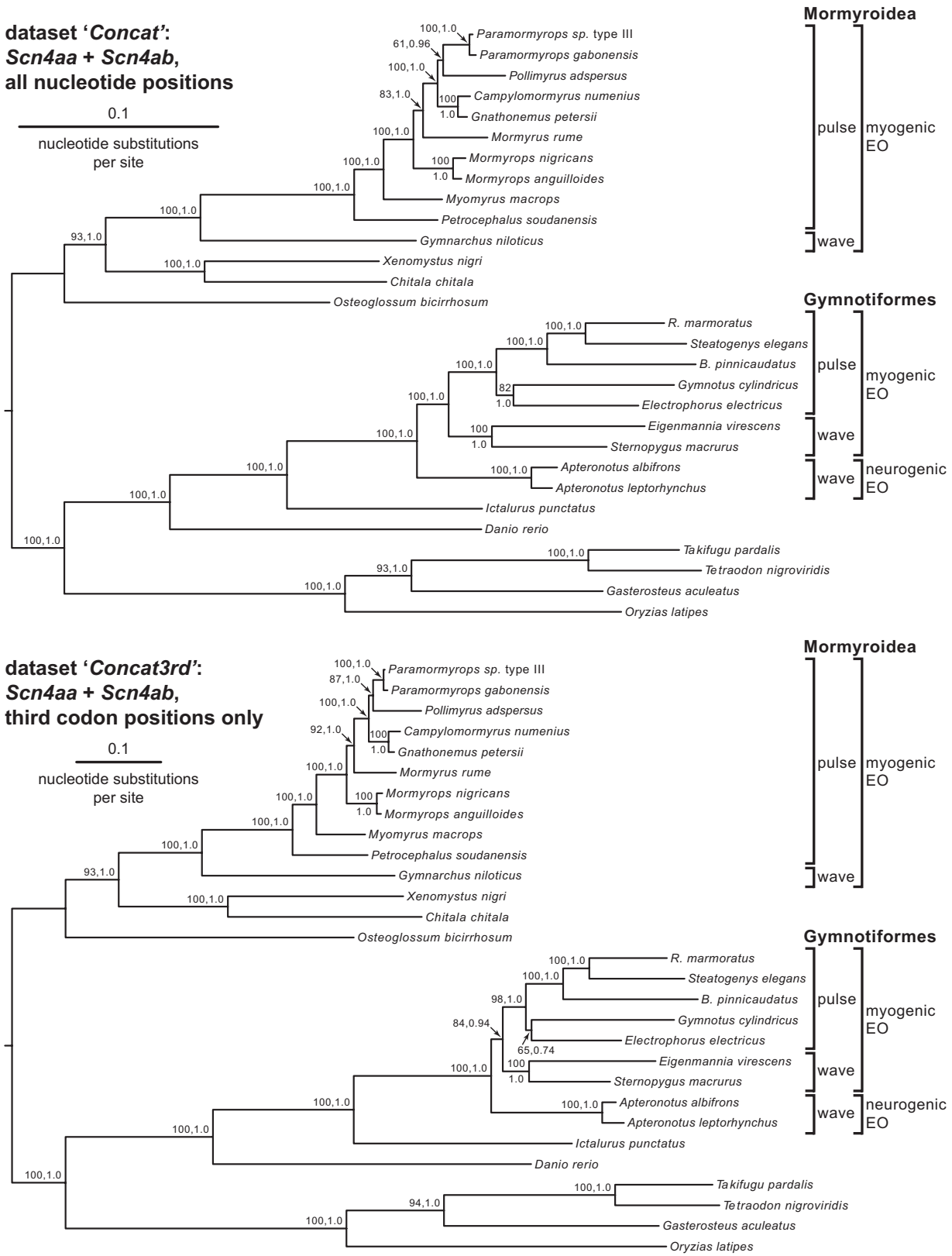

Fig. S2. Estimated phylogenetic trees of teleost fishes based on concatenated data combining both paralogs, *Snc4aa* and *Scn4ab*. The *Upper* tree is based on the dataset *Concat*, which includes all nucleotide positions. The *Lower* tree is based on the dataset *Concat3rd*, which includes only third-codon-position nucleotides. Support values are given at each node: bootstrap proportion (%), followed by Bayesian posterior probability.

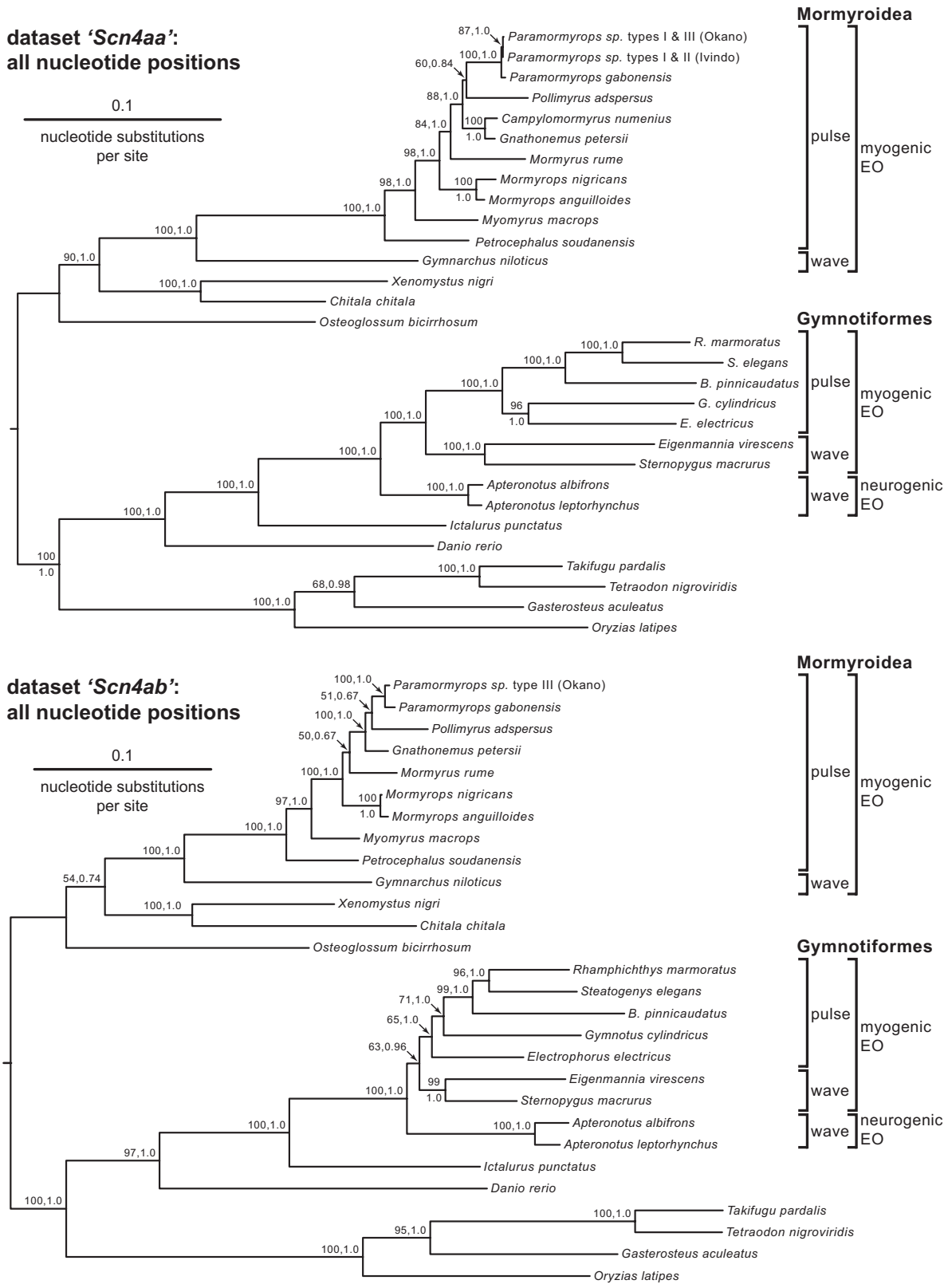

Fig. S3. Estimated phylogenetic trees of teleost fishes based on each individual paralog (*Upper*: dataset *Scn4aa*; *Lower*: dataset *Scn4ab*). Both datasets include all nucleotide positions (i.e., first-, second-, and third-codon positions). Support values are given at each node: bootstrap proportion (%), followed by Bayesian posterior probability.

Fig. S4. (Continued)

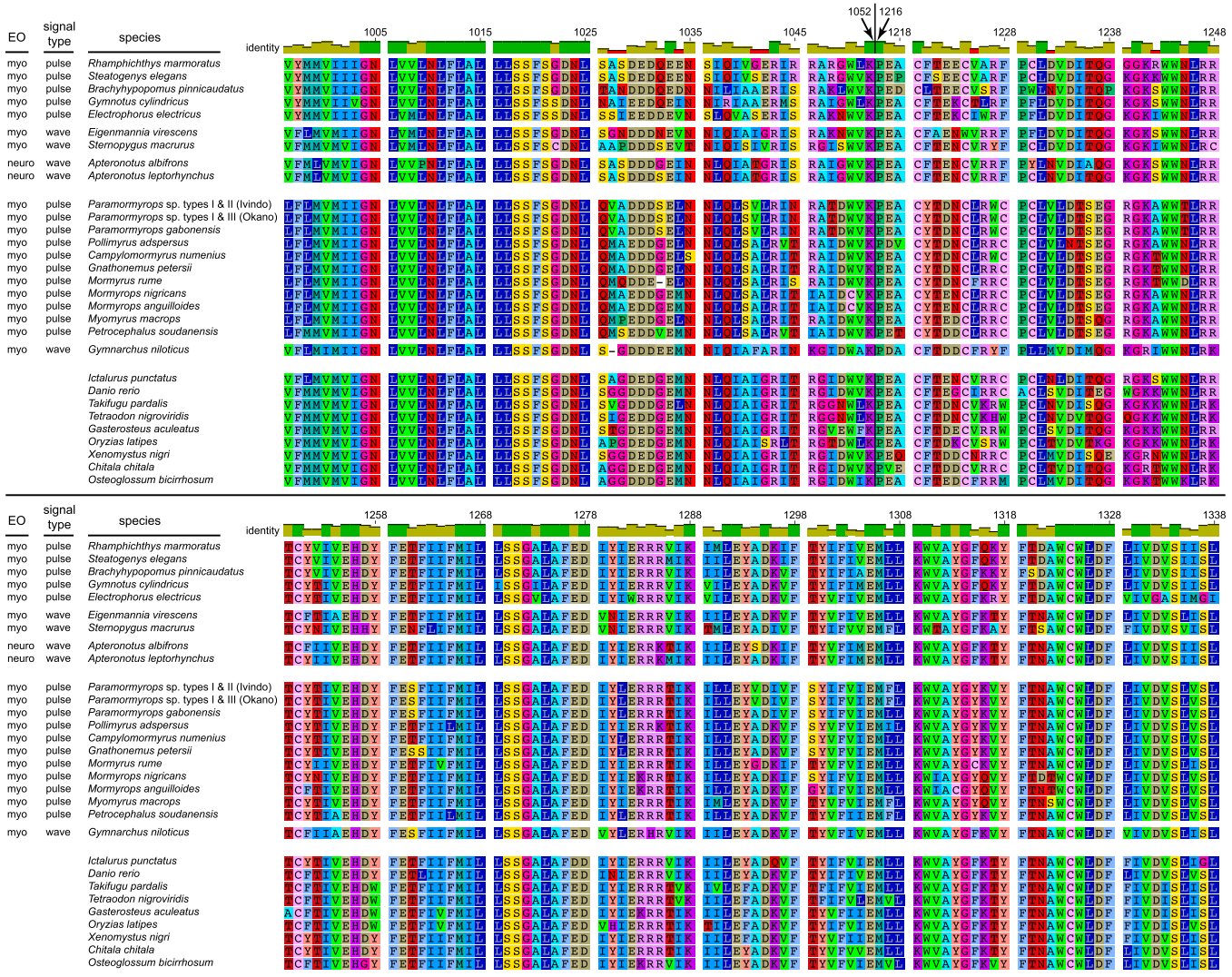

Fig. S4. (Continued)

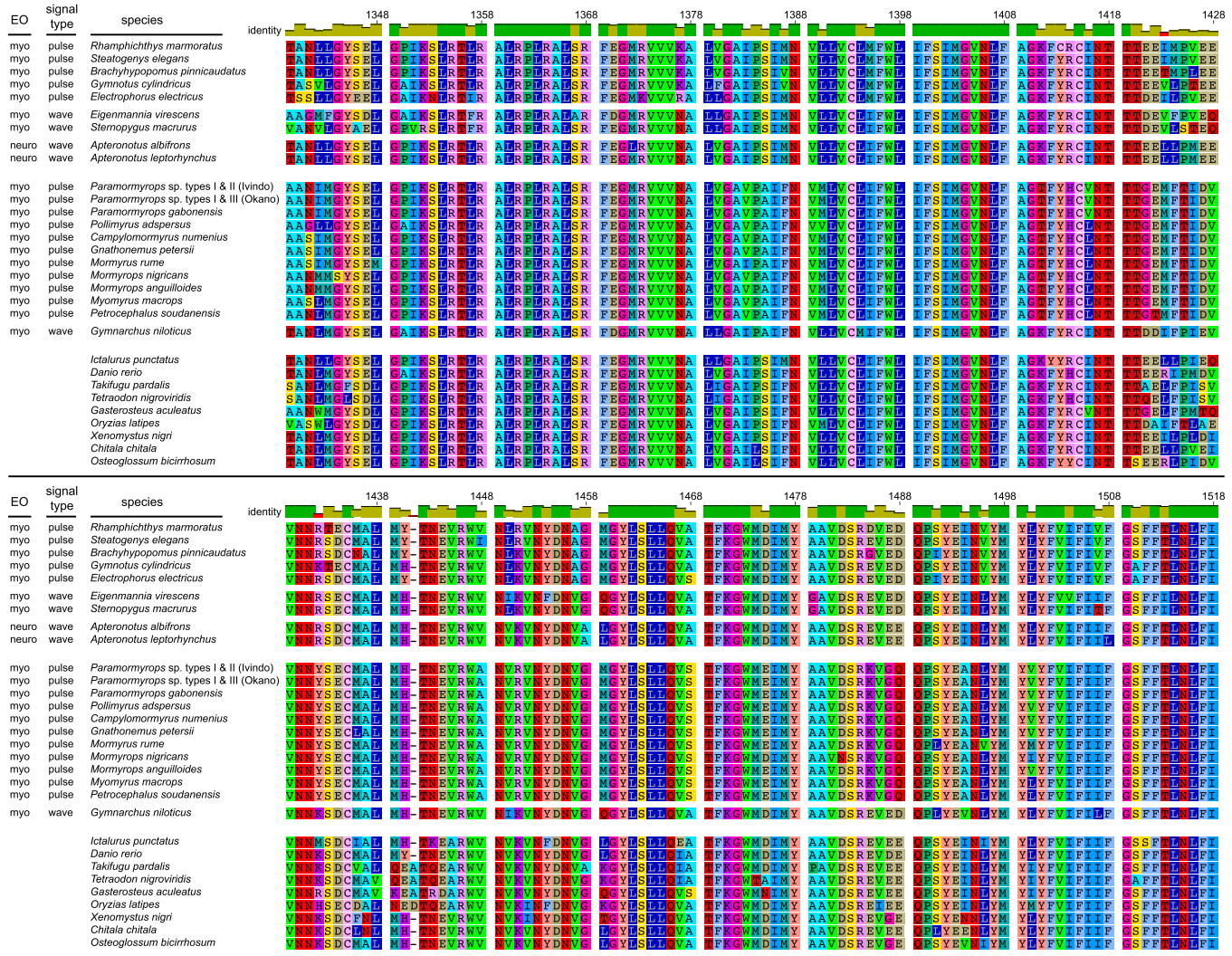

Fig. S4. (Continued)

Fig. S4. (Continued)

Fig. S4. Complete amino acid alignment for the investigated regions of $Na_v1.4a$, the voltage-gated Na^+ channel that in electric fish is expressed in myogenic electric organ but not skeletal muscle. The alignment spans five pages; each page shows two contiguous portions of the Na^+ channel (top and bottom of the page). Species names are given at the left of each portion, with electric organ (EO) and electric signal types given at the far left for electrogenic fishes. Hypothesized natural groupings of electric fishes (e.g., based on our phylogenetic results for gymnotiforms) are separated by empty space. The alignment was generated as described in *Sequence Alignments*, and it was formatted using Geneious version 4.8 (Biomatters). Amino acid residues are indicated by the standard letter code and MacClade color scheme (45). Residue numbers at the top of each alignment correspond to the standard electric eel reference sequence (16). Sequence identity (the default plot generated by Geneious) is also shown along the top: green, residue at that position invariant across species; yellow, less than complete identity; red, high amino acid variability at that position. A thin vertical line between concatenated residues 1052 and 1216 at the top of the second page of the figure (see arrows) indicates the excluded hyper-variable sequence.

Table S1. Fish species (and morphs) for which cDNA sequence data were collected for this study

Species	Source	CUMV catalog no.	*Scn4aa* ($Na_v1.4a$) GenBank accession no.	*Scn4aa* ($Na_v1.4a$) Sequence length (bp)	*Scn4ab* ($Na_v1.4b$) GenBank accession no.	*Scn4ab* ($Na_v1.4b$) Sequence length (bp)
Mormyroid electric fishes						
Campylomormyrus numenius	Aquarium trade (wild caught); 5736*	CU 91892	GU362014	2901	†	†
Gnathonemus petersii	Aquarium trade (wild caught)		GU362015	2901	GU362039	2805
Paramormyrops sp. type I morph	Okano River, Gabon; 6401*	CU 92249	GU362016	2901	†	†
Paramormyrops sp. type III morph	Okano River, Gabon; 6450*	CU 92252	GU362017	2901	GU362040	2799
Paramormyrops sp. type I morph	Ivindo River, Gabon; 6227*	CU 92331	GU362018	2901	†	†
Paramormyrops sp. type II morph	Ivindo River, Gabon; 6245*	CU 92330	GU362019	2901	†	†
Paramormyrops gabonensis	Ivindo River, Gabon; 6269*	CU 92344	GU362020	2901	GU362041	1788
Pollimyrus adspersus	Aquarium trade (wild caught); 5726*	CU 91329	GU362021	2901	GU362042	2847
Mormyrus rume	Aquarium trade (wild caught)		GU362022	2757	GU362043	2871
Mormyrops nigricans	Lékoli River, Congo; 6163*	CU 92985	GU362023	2901	GU362044	1641
Mormyrops anguilloides	Lékoli River, Congo; 6200*		GU362024	2871	GU362045	2751
Myomyrus macrops	Lékoli River, Congo; 6182*	CU 92394	GU362025	2901	GU362046	2874
Petrocephalus soudanensis	Aquarium trade (wild caught); 5727*	CU 91327	GU362026	2886	GU362047	2880
Gymnarchus niloticus	Aquarium trade (wild caught)		GU362027	2910	GU362048	2856
Gymnotiform electric fishes						
Rhamphichthys marmoratus	Aquarium trade (wild caught)		GU362028	2736	GU362049	2823
Brachyhypopomus pinnicaudatus	Aquarium trade (wild caught)		GU362029	2832	GU362050	2292
Steatogenys elegans	Aquarium trade (wild caught)		GU362030	2802	GU362051	2838
Gymnotus cylindricus	Aquarium trade (wild caught)		GU362031	2760	GU362052	2751
Electrophorus electricus	Aquarium trade (wild caught)		‡	‡	GU362053	2718
Eigenmannia virescens	Aquarium trade (wild caught)		GU362032	2814	GU362054	2817
Apteronotus albifrons	Aquarium trade (wild caught)		GU362033	2802	GU362055	2790
Apteronotus leptorhynchus	Aquarium trade (wild caught)		GU362034	2781	GU362056	2709
Nonelectric teleost fishes						
Xenomystus nigri	Aquarium trade (wild caught)		GU362035	2892	GU362057	2796
Chitala chitala	Aquarium trade (wild caught)		GU362036	2922	GU362058	2640
Osteoglossum bicirrhosum	Aquarium trade (wild caught)		GU362037	2913	GU362059	2805
Ictalurus punctatus	Fish farm (Austin, TX)		GU362038	2826	GU362060	2073
Takifugu pardalis	Aquarium trade (wild caught)		‡	‡	GU362061	2781

GenBank accession numbers are provided for all of our sequence data, which we have made publicly available through the National Center for Biotechnology Information. CUMV, Cornell University Museum of Vertebrates.

*Assigned specimen number.

†mRNA transcript neither sequenced nor available elsewhere.

‡Analyzed transcript taken from previously published sequence data (see *cDNA Cloning and Sequencing*).

Table S2. Primers and cycling conditions for *Scn4aa* expression assays

Species (fragment size in bp)	RT primer	Upper PCR primer	Lower PCR primer	Temperature (°C)	No. of cycles	No. of introns
Nonelectric ostariophysans						
Danio rerio (550 bp)	TTCAGGCTCAGGTTCTGGAG	AGGACTGCGTTTGTAAGATCTCTGA	ATTAAGAGTGGCCTTTGCGTCTTCC	56.6	30	2
Ictalurus punctatus (765 bp)	GCTTCTTTAGTGTGCATTAGTG	GTCAGATGATTTATCAGAAGAAGCGA	TGTTCACTTGCTCGATGGGTAAAAGT	55.8	30	4
Electrogenic ostariophysans						
Apteronotus albifrons (481 bp)	CTCCAAGATGATCTTAATCGTCT	CGGTGACAATCTCTCGGCAAGCGACG	TTGAGATAAGGAAATCGCCTCACACA	58.3	30	2
Gymnotus cylindricus (726 bp)	TGTTCACTTCCTCCGTGGGCAGA	CAACGCAATTGAAGAAGACCAGGAG	TAACATCTCCATTATGAAAATATAT	52.9	30	3
Steatogenys elegans (465 bp)	TCTTAATCATCCTTCGACGCTCG	TTCATTCAGCGGTGACAATCTCAGTG	CTTCCGAAAAGCAGGGTTCTGGATCA	58.9	30	1
Rhamphichthys marmoratus (465 bp)	GGCGTCAGTGAAGTACTTTTGAAAC	CGAAGACCAAGAGGAAAACAGTA	CATCCAGACAAGGAAATCGTGCC	56.4	30	2
Nonelectric osteoglossomorphs						
Osteoglossum bicirrhosum (985 bp)	CTTTTCCTCCTAACTTTTTC	GCCATTGAAGCCCCCAGTGCAGATC	ACCTCATTGGTGTGCATTAGGGCCA	60.4	30	5
Xenomystus nigri (1126 bp)	ACTGCTCAGGCTCCACCAGCTCT	TGTGCCCAAGTTCGTCATCC	TCCTCTTTACCATTTTTCTC	53.7	30	4
Electrogenic osteoglossomorphs						
Gymnarchus niloticus (785 bp)	AACTCGCTGGGCTGAATGTC	CCAGAGTTCGAACATGTGCTCAGTG	CATGGTAAAGGCCTCGATGTTGGGG	58.5	30	3
Petrocephalus soudanensis (537 bp)	AATCTCCTCCCATCTTGTTC	GCAAATGTCAGAAGATGATG	CCAGCACCAGACAAGGACAGCGT	57.5	30	2
Myomyrus macrops (837 bp)	CTCATATGAGGGCTGTTGAC	AATGCTTGATACTGATGGGGAGGATT	CATTTGTGTGCATGAGGGCCATACAC	58.7	30	5
Percomorph						
Gasterosteus aculeatus (1073 bp)	CGTCCCGAGTGGCCTCCTTG	ATCCTGTGTGGGGAGTGGATCGAGACC	TCCCATCCAGTTGGCAGCCAAACTA	60.2	30	5

Temperature and no. of cycles are the PCR annealing temperature and the number of cycles that yielded a strong signal without saturation of product. Resulting fragment size, traversing the number of introns listed at the far right, are given in parentheses next to the species name (shown at the far left). All data correspond to the results shown in Fig. 3.

Table S3. Primers and cycling conditions for *Scn4ab* expression assays

Species (fragment size in bp)	RT primer	Upper PCR primer	Lower PCR primer	Temperature (°C)	No. of cycles	No. of introns
Nonelectric ostariophysans						
Danio rerio (538 bp)	CTTCCTCTTCTACAAGATCTTCTAC	CAGACTCTGTGTGTAAAATCTCATC	TTCATCTCCATTTTTCTTTTTCATA	54.5	30	2
Ictalurus punctatus (720 bp)	TTCATCCATACCGGAATTGATC	ATTGATGAAAAAAAGAAACTGAACGA	GCTTGGAAATACTCTTCAGAGGTATA	56.1	30	4
Electrogenic ostariophysans						
Apteronotus albifrons (511 bp)	CTCAAGTATGACCTTTATTGCTC	ATCAGATGGTGAAATGAACAATC	GTCTTGCGGAAATTCCACAATAA	54.5	30	2
Gymnotus cylindricus (695 bp)	GTACAGCTGTATCATTGTTTTCC	CATTGTCAGAATCACTAGAGGTACT	AAGTAAACCTTGAATCCATAGGCTG	54.8	30	3
Steatogenys elegans (422 bp)	TGCATTTCTCTGTGTAACAGTCC	TCAGATGATGATGGTGAAATGAACAA	GGCTGTCAGGATCTTCTTCTTCGTCT	56.6	30	1
Rhamphichthys marmoratus (481 bp)	CCATGCATTTGTGAAGTAAACCTTG	ATCAGATGATGATGGTGAAATGA	CATGGACATCGCCTGATGCATTT	55.3	30	2
Nonelectric osteoglossomorphs						
Osteoglossum bicirrhosum (1017 bp)	ACTCTCATAAAAAGGCTGTT	GGTCCAGATCATCAGGCAGAYCTTG	GGTGTTGTTCTCCCGCATTAGATTG	58.1	30	5
Xenomystus nigri (1163 bp)	GTTCCACCAGGACCGCCCTCTGC	TCTGTCTGATGTAGAGATGT	AGCAGGTTTGTCACGGGCCA	53.2	30	4
Electrogenic osteoglossomorphs						
Gymnarchus niloticus (804 bp)	TTAAGACGAAAACATCCTTC	GGTCTTTAGAAAATGGGTGTACTTC	AACTATAGAGGCCTTGACAAAATTG	56.6	30	3
Petrocephalus soudanensis (490 bp)	CCTGTCCTCCAAACTTTGAT	ATCGCTATCAACAGGATCAC	GTCCACCATTAGGAAGGGGAAGC	55.7	30	2
Myomyrus macrops (806 bp)	ACGTTGAAGATGGAGGGTAT	GAGAGAACAACCTCCAGATCGCTATCA	GCTGATCAAAGAAACATCCACAATGAG	57.0	30	4
Percomorph						
Gasterosteus aculeatus (1099 bp)	CTCCGTGAAATTTGCCTTTATC	ACGTTCAGGGCCTGTCTGTGCTGCGCT	GTACTCCAGGATCGTCTTTATGACC	60.2	30	5

Temperature and no. of cycles are the PCR annealing temperature and the number of cycles that yielded a strong signal without saturation of product. Resulting fragment size, traversing the number of introns listed at the far right, are given in parentheses next to the species name (shown at the far left). All data correspond to the results shown in Fig. 3.

Table S4. Rankings and scores of the model-partitioning schemes applied to the four alignments

Alignment	Partition	*ln*L	No. of parameters	AIC
Scn4ab	(1) (2) (3)	−22698.82	30	45457.64
	(12) (3)	−22817.84	19	45673.68
	(123)	−23524.92	10	47069.84
Scn4aa	(1) (2) (3)	−24537.93	31	49137.85
	(12) (3)	−24639.11	19	49316.22
	(123)	−25353.78	10	50727.55
Concat	(11) (2) (2) (33)	−47384.54	42	94853.08
	(1) (2) (1) (2) (33)	−47375.30	53	94856.60
	(1) (2) (3) (1) (2) (3)	−47373.34	62	94870.69
	(11) (22) (33)	−47408.21	31	94878.43
	(1212) (33)	−47614.47	19	95266.94
	(123123)	−49038.37	10	98096.75
	(123) (123)	−49031.42	21	98104.83
Concat3rd	(33)	−27433.73	9	54885.46
	(3) (3)	−27431.50	19	54901.00

Table S5. Models chosen by MODELTEST using the AIC criterion for various subsets of the alignments

	Subset of alignment				
Alignment	Full	First position	Second position	Third position	First position + second position
Scn4ab	GTRIG	GTRIG	TVMIG	TVMIG	TVMIG
Scn4aa	GTRIG	GTRIG	GTRIG	TVMIG	TVMIG
Concat	GTRIG	GTRIG	GTRIG	TVMIG	TVMIG