Supplementary Materials for

Genomic basis for the convergent evolution of electric organs

Jason R. Gallant, Lindsay L. Traeger, Jeremy D. Volkening, Howell Moffett, Po-Hao Chen, Carl D. Novina, George N. Phillips Jr., Rene Anand, Gregg B. Wells, Matthew Pinch, Robert Güth, Graciela A. Unguez, James S. Albert, Harold H. Zakon,* Manoj P. Samanta,* Michael R. Sussman*

*Corresponding author. E-mail: msussman@wisc.edu (M.R.S.); manoj.samanta@systemix.org (M.P.S.); h.zakon@austin.utexas.edu (H.H.Z.)

Published 27 June 2014, *Science* **344**, 1522 (2014)
DOI: 10.1126/science.1254432

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/344/6191/1522/suppl/DC1)

Supporting Online Materials

Supplemental Online Materials and Methods

1. Animal Sources

Electrophorus electricus (*26*) - In all, three fish were used in this study, and purchased from a tropical fish dealer, Tri-County Tropicals (Richmond Hill, NY). Fish approximately 70 cm in length were housed either individually, or with one other electric eel in aerated aquaria maintained at 26-28°C. From the first animal, the main EO was used for gDNA sequencing. From the second animal, whole brain, spinal cord, whole heart, skeletal muscle, Sachs' EO, main EO, and Hunter's EO were isolated and used for RNA sequencing. From the third animal, whole kidney was isolated and used for RNA sequencing.

Eigenmannia virescens (27, 28) - One *Eigenmannia* was purchased from a tropical fish dealer (Segrest Aquarium) and was used for this study.

Sternopygus macrurus (29) - A fresh-water species of knifefish native to South America was obtained commercially from Ornamental Fish (Miami, FL). Adult fish 30-50 cm in length were housed individually in 15- to 20-gallon aerated aquaria maintained at 25–28°C and fed three times weekly. One fish of undetermined sex was used in this study.

Malapterurus electricus (*30*) - A freshwater species of catfish, native to Africa, was commercially obtained from the Route 4 Aquarium in Elmwood Park (Elmwood Park, NJ). Adult fish were housed in groups in aerated aquaria maintained at 25-28˚C and fed daily. We dissected electric organ and muscle from two specimens.

Brienomyrus brachyistius (*31, 32*) – A freshwater mormyrid species native to central Africa and bred in the laboratory by JRG from two parents obtained commercially from Baileys Wholesale Tropical Fish (San Diego CA) was used for this study. Adult fish were housed in groups in aerated aquaria maintained at 25-28˚C and fed three times weekly. We performed RNA sequencing (one for SM and one for EO) on specimens of undetermined sex.

All procedures used followed the American Physiological Society Animal Care Guidelines, and were approved by the Institutional Animal Care and Use Committee at Cornell University, New Mexico State University, University of Texas at Austin, Michigan State University and the University of Wisconsin-Madison.

2. Genome Sequencing of *Electrophorus electricus*

Genomic DNA Isolation

Genomic DNA was isolated from the main EO using phenol-chloroform extraction (*33*) with several modifications. Tissue from the main electric organ was pulverized in liquid nitrogen using a mortar and pestle, and 1 ml solution D was added per 100 mg

tissue. 500 ul aliquots of sample were loaded into 1.5 ml microcentrifuge tubes. To each of these tubes, 500 ul Tris-saturated phenol (pH 7.8), 500 ul chloroform-IAA, and 100-200 ul nuclease-free water were added. Tubes were mixed by inversion and spun in a microcentrifuge at 14,000 rpm for 5 minutes. The aqueous phase was transferred to a new tube and DNA was precipitated by addition of 1ml 100% ethanol and 100 ul 3M sodium acetate and incubation at -20°C for approximately 1 hour. DNA was pelleted using a microcentrifuge at 14,000 rpm for 10 minutes and the supernatant was discarded. DNA was resuspended in DEPC-treated water and put into a 42°C water bath for approx. 10 minutes until dissolved. 0.5 ul RNaseA (1 mg/ml) was added, and samples were incubated at room temperature for 20 minutes. Samples were subjected to a second round of phenol-chloroform extraction as described above. DNA was precipitated from the aqueous phase with EtOH/ NaAC and pelleted as described above, and pellets were dried slightly (2-4 minutes) in a sterile hood. DNA was resuspended in 50-100 ul DEPC-treated water, and tubes were combined. Sample concentration was measured on a Nanodrop spectrophotometer (Thermo Scientific) and the integrity of the DNA and absence of RNA were evaluated by running samples on a 1% agarose gel prior to submission to the University of Wisconsin-Madison Next Generation Sequencing Facility.

Library Preparation and DNA Sequencing

Genomic DNA libraries were sequenced on various platforms as summarized in Table S5.

Illumina GAIIx – Paired-end gDNA libraries were prepared using the Illumina Paired End Sample Preparation Kit following manufacturer's guidelines, with one modification: gDNA was size-selected on an Invitrogen E-gel instead of a standard agarose gel.

Illumina HiSeq 2000 – Paired-end (PE) DNA libraries were prepared using the Illumina TruSeq DNA Sample Preparation Kit following manufacturer's guidelines, with one modification: gDNA was size-selected on an Invitrogen E-gel rather than a standard agarose gel.

Illumina HiSeq 2000 – 2-5kb mate-pair libraries were prepared using the Illumina Mate Pair Library Preparation Kit v2 following manufacturer's guidelines.

3. RNA Extraction and Library Preparation for RNA-Seq in *Electrophorus electricus, Sternopygus macrurus, Eigenmannia virescens, Malapterurus electricus* and *Brienomyrus brachyistius*

Various mRNA libraries were sequenced on various platforms as summarized in Table S6.

E. electricus - Total RNA extraction was performed for each of the eight tissues following the phenol/chloroform extraction method outlined in Chomczynski and Sacchi

2006 (*33*), with the following modifications. Tissue samples were ground in liquid nitrogen using a ceramic mortar and pestle. Following grinding, 1 ml per 100 mg tissue of solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0; 0.5%(wt/vol) *N*-lauroylsarcosine; 0.1 M 2-mercaptoethanol) was added, mixed, and transferred to 2 ml microcentrifuge tubes containing a single 50 mm steel bead. Tissue was homogenized on a bead mill for 2 minutes at 20 Hz at 4°C, tubes were rotated, and homogenized again for 2 minutes at 2 Hz. Tubes were centrifuged at 14,000 RPM for 10 minutes at 4°C, and supernatant transferred to fresh nuclease-free microcentrifuge tubes, 500 ul/tube, for phenol/chloroform extraction. Following the first extraction and precipitation, RNA pellets were resuspended in DEPC-treated water and DNase treated following Qiagen MinElute protocol in appendix C, with DNaseI. A second phenol/chloroform extraction was performed, followed by a second ethanol precipitation, and RNA wash step as described in Chomczynski and Sacchi (*33*). cDNA libraries were constructed from purified RNA using the Illumina TruSeq RNA Sample Preparation (v.2) kit (San Diego, CA). Libraries were sequenced on an Illumina HiSeq 2000 using 100bp single-end reads (1x100bp).

S. macrurus - Fish were anesthetized with 2-phenoxyethanol (1.0 mL/L) to excise ventral skeletal muscle and caudal EO under a dissecting microscope. Tissues were blotted dry, weighed, and immediately flash frozen in liquid nitrogen. Total cellular RNA was isolated from both tissues by chopping the frozen tissues into smaller pieces and subsequently pulverizing the tissues in liquid nitrogen using a mortar and pestle. Pulverized tissues were re-suspended in TRIzol reagent (Invitrogen, Carlsbad, CA) and total RNA extracted following manufacturer's instructions. To remove residual DNA, total RNA was treated with DNase I, Amplification Grade (Invitrogen) according to manufacturer's instructions. Total RNA was then purified using phenol:chloroform:isoamyl alcohol extraction followed by isopropanol precipitation. cDNA libraries were constructed from purified RNA using the Illumina TruSeq RNA Sample Preparation (v.2) kit (San Diego, CA). Libraries were sequenced on an Illumina HiSeq 2000 using 100bp paired-end reads (2x100bp).

E. virescens – EOs were dissected by cutting the tail and removal of the skin. A piece of muscle was dissected from the back of the fish. From each sample, total RNA was extracted with RNA STAT-60 and then was treated with a procedure to remove ribosomal RNA. Libraries were sequenced on an Illumina HiSeq 2000 using 99bp paired-end reads (2x99bp).

M. electricus -Fish were euthanized by overdose of MS-222. EOs were dissected by removal of the skin and electric organ which are closely attached over the majority of the body surface. Skin and electric organ were separated using fine forceps. Trunk skeletal muscle was then dissected from directly below the sites where electric organ were dissected. This region was approximately ~2×1×0.5 cm, caudal to operculum, dorsal to lateral line. Tissue was stored in RNA*later*, and total RNA was extracted using an RNeasy fibrous tissue extraction kit (Qiagen, Inc.). Samples were submitted to the Michigan State University RTSF for RNA sequencing using the TrueSeq mRNA sample preparation kit (Illumina, Inc.) Resulting libraries were sequenced on an Illumina HiSeq 2500 using 150 bp paired-end reads (2x100bp).

B. brachyistius - Fish were euthanized by overdose of MS-222. EOs were dissected by removal of the skin and muscle from the caudal peduncle, excision of the EO and spinal column, and finally removal of the spinal cord by inserting a fine pin into the vertebral column. Trunk skeletal muscle was dissected from ~2×1×0.5 cm, caudal to operculum, dorsal to lateral line; skin removed. Tissue was immediately frozen in liquid nitrogen, then pulverized using pestle and mortar, and total RNA was extracted using TRIzol solution (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Total RNA was PolyA+ purified using a FastTrack MAG mRNA isolation kit (Invitrogen). cDNA libraries were constructed from purified mRNA using the NEBnext Sample Kit for Illumina Sequencing (New England Biolabs). Libraries were sequenced on an Illumina HiSeq 2000 using 100 bp paired-end reads (2x100bp).

4. Draft Genome Assembly of *Electrophorus electricus*

Estimation of Genome Size

The haploid genome size of *E. electricus* has not been measured empirically, but the sizes for other species in the class Gymnotiformes are available (*31, 32, 34-36*). In order to obtain an empirical value of genome size based on sequence data, we ran the 'preqc' module (-preqc) of SGA assembler on error-corrected PE reads (*37*). The module gave an estimate of ~720 Mb by splitting the PE reads into 31-mers and analyzing their distribution. The completion of the SGA assembly (specifying the parameter OD=75, in addition to default parameters) using PE reads produced a genome of size 533 Mb after exclusion of small contigs (<2x read length). The differences in genome size estimation highlights the uncertainty of genome size predictions and methods. However, these estimated sizes for *E. electricus* genome agree well with the relationship between genome size and GC content previously determined in teleost fishes (*38*).

We also split the PE libraries from both strands into all possible 21-mers and evaluated their distribution. Distribution of 21-mers showed a single peak at frequency=55 (ignoring the error peak at frequency=1). This suggests that the PE reads sampled the genome at 55x k-mer coverage.

Assembly of Genomic DNA

A draft genome assembly was built from the Illumina paired end and mate pair reads using SOAPdenovo2, a de Bruijn graph-based genome assembler (*39*). Reads were assembled with de Bruijn graph parameter k=47, a value that produced the best N50 scaffold size. All analyses in the paper were performed using this version of the assembly. The assembled genome build is summarized in Table S7.

Transcriptome mapping-based evaluation

Because our genomic DNA was isolated from adult, differentiated tissue (main EO), we wanted to confirm the suitability of our SOAPdenovo2 genome build as a reference for *E. electricus* as a whole. To do this, we mapped the independent transcriptome assembly generated from RNA sequencing reads from eight *E. electricus* tissues (see Supplementary Materials, section 6) to the SOAPdenovo2 build using GMAP (*40*) We found that out of the 365,443 transcripts in our eight-tissue Trinity assembly, 357,859 (97.92%) transcripts aligned to the SOAPdenovo2 build. This high degree of overlap between the two independent assemblies is evidence that our SOAPdenovo2 genome build is suitable for subsequent genome and gene analyses.

Analysis of Assembly Quality of the SOAPdenovo2 Build

CEGMA-based evaluation: The SOAPdenovo2 genome build was analyzed using CEGMA v2.4 (*41*) with default parameters in order to evaluate completeness of coding regions. CEGMA located 446 (97%) of the 458 genes included in its core set. Of the subset of 248 most-conserved genes defined by CEGMA, 217 (88%) were identified as full-length and 245 (99%) were identified as either full or partial.

K-mer based evaluation: We also evaluated the SOAPdenovo2 assembly for completeness by checking what fraction of 21-mers within the PE reads were incorporated into the assembled genome, and the number of times those 21-mers were present in the assembly. Earlier we described the 21-mer distribution with a peak at frequency=55 representing the k-mer coverage of the PE reads. Therefore, it is expected that the 21-mers present with frequency ~110 in PE library are, on average, present twice in the actual genome, frequency ~165 are, on average, present thrice and so on. Hence, a comparison between 21-mer distribution of the PE reads and 21-mer distribution of assembled genome should illustrate the level of completeness of the assembled genome. Moreover, it highlights whether the missing regions are from the non-repetitive or repetitive parts of the actual genome.

The comparison between 21-mers from SOAPdenovo2 assembly and the PE reads suggests that the assembly is over 98% complete in the non-repetitive regions represented by 21-mer frequency between 40 and 80 in PE reads. The distribution tapered off below frequency of 40. The multiplicity of 21-mers in the genome increased for those with PE frequency> 100, but the rate of increase was lower than expected. This suggests that the SOAPdenovo2 assembly provides sub-optimal assembly of the repetitive and low-coverage regions.

5. Gene Annotation in Electrophorus electricus

Genome structural annotation

RNA-Seq reads from eight *E. electricus* mRNA tissue libraries were mapped to the genome using TopHat v2.0.4 (*42*) and specifying "--microexon-search -i 20 -I 50000" in addition to the default parameters. For the purpose of gene prediction, alignments from all eight tissues were merged and extrinsic hints files for "intron" and "exonpart" features

were generated for input into AUGUSTUS v2.6 (*43*). Protein-coding gene models were predicted in all genome scaffolds at least 500 bp in length with AUGUSTUS, using human-specific parameters ("--species=human") and *E. electricus* extrinsic transcriptional evidence, with the following altered parameters in the otherwise default extrinsic.M.RM.E.W.cfg file:

```
exonpart        1     .997  M    1  1e+100  RM  1     1    E 1    1e2  W 1  1.007
    exon        1       1   M    1  1e+100  RM  1     1    E 1    1e4  W 1    1
  intron        1       .3  M    1  1e+100  RM  1     1    E 1    1e6  W 1    1
 UTRpart        1    1 .96  M    1  1e+100  RM  1     1    E 1    1    W 1    1
```

Additionally, "--alternatives-from-evidence=true --allow_hinted_splicesites=atac --UTR=on" was specified. Based on an initial manual inspection, CDS features predicted from AUGUSTUS were refined to use the first in-frame start codon and to choose the longest ORF except when homology to *D. rerio* (blastp (*44*) against Ensembl Zv9 build 69 (*45*) e-value <= $1E^{-20}$) suggested otherwise.

Functional annotation of predicted gene models

Gene symbols and functional annotations were assigned to AUGUSTUS gene models by comparison to D. rerio Zv9 build 70 protein sequences. D. rerio protein sequences were used to search the E. electricus AUGUSTUS predicted protein sequences at an e-value cutoff of $1E^{-10}$. Results were compiled based on D. rerio gene ID and E. electricus hits were ordered based on alignment bit score. In addition, the top-scoring D. rerio genes for each E. electricus database entry were tracked and ranked. Thus, each D. rerio query was associated with an ordered ranking of E. electricus hits and each E. electricus database entry was associated with an ordered ranking of D. rerio queries. In the initial round of assignments, reciprocal best-hit pairs were identified and recorded and the E. electricus gene was assigned an annotation class of 'reciprocal'. In addition, for each D. rerio query, lower-scoring *E. electricus* hits for which the D. rerio gene was the top query match were checked in an iterative fashion for both adjacency to the top-scoring E. electricus gene and alignments to the D. rerio gene that overlapped by <20%. If identified, these *E. electricus* gene models were assigned to the same *D. rerio* gene and given an annotation class of 'split' indicating that they are likely to be incorrectly split from the primary gene model. This process was repeated for each locus until no additional assignments were made.

In a second round of annotation, all unannotated *E. electricus* genes that were positioned within 5000bp of the end of a scaffold were checked against the currently assigned genes sharing the top *D. rerio* query. If both fell at the end of a scaffold and *D. rerio* alignments overlapped by <20% they were assigned to the same *D. rerio* gene and give an annotation class of 'split_scaff', indicating that they likely belong to the primary gene model but are split across scaffolds. In the third round of annotation, *D. rerio* genes without assigned *E. electricus* genes were checked and if remaining unassigned matches were found that agreed by synteny with assigned *E. electricus* genes for immediately adjacent *D. rerio* genes, they were assigned to that *D. rerio* gene and given an annotation

class of 'synteny'. All remaining unassigned *E. electricus* genes which had any matches to *D. rerio* queries were assigned to the top *D. rerio* query and given an annotation class of 'secondary' to indicate the increased uncertainty in the assignment. After automated annotation, genes which were examined by hand in the course of work and had their annotations adjusted were assigned an annotation class of 'manual'.

All AUGUSTUS genes without high-scoring *D. rerio* matches were subsequently searched against the GenBank non-redundant protein database using blastp (*44*) with an e-value cutoff of $1E^{-10}$. Genes with GenBank hits were assigned to the top hit with a match class of 'nr'. Taking all together, the coding content of the genome was calculated as 22,228, or approximately 22,000, protein-coding genes with homology to published sequences (Table S1).

6. Transcriptome Assembly

For each fish, short read libraries from different tissues were combined and quality control and filtering was performed using the fastx toolkit (CSHL) as well as scythe and sickle (UC-Davis). The transcriptome assembly pipeline Trinity (*46*) (r2012-06-08) was utilized to perform de-novo transcript assembly of short reads for each fish using default parameters.

E. electricus RNA-Seq libraries were assembled in three phases with accumulation of additional experimental data.

i) All reads from three initial GA2 RNA-Seq libraries (muscle, main EO, Sachs' EO) were combined together and assembled using Trinity (default parameters).

ii) Reads from seven HiSeq tissues (brain, spinal cord, heart, muscle and three different EOs) were combined together and assembled using Trinity (default parameters). The assembly steps (i) and (ii) resulted in ~449,000 assembled transcripts.

iii) After completion of RNA-Seq experiment on kidney (conducted separately), reads from all eight *E. electricus* tissues were combined and assembled using Trinity, specifying --kmer_method meryl. This resulted in ~365,000 assembled transcripts.

E. virescens The reads from the *E. virescens* RNA-Seq experiment on EO and skeletal muscle were combined together. Execution of Trinity on the combined library was slow due to the size of the library. Therefore, we used two strategies - (i) subset of reads picked randomly from the library, (ii) digital normalization (http://ged.msu.edu/angus/diginorm-2012/tutorial.html) - to reduce the number of reads, and then performed Trinity assembly of the reduced library using default parameters. All downstream evaluations were performed based on the combined set of contigs generated from these assemblies.

S. macrurus - The *S. macrurus* transcriptome assembly contained 326,623 sequence contigs representing 221,914 subcomponents. Of these sequences, 163,477 were at least 500 bp in length and 63,408 were at least 2,000 bp in length. The average sequence length was 1287 bp.

M. electricus - Trinity assemblies were generated from SM/EO paired end reads using default parameters with the Trinity 2013 Nov 10th release. The assembly contained 181,633 transcript contigs.

B. brachyistius - Trinity assemblies were generated from SM/EO paired end reads using default parameters with the Trinity 2012 October 5th release. The assembly contained 147,923 transcript contigs.

7. Calculation of Expression Levels

E. electricus - Tissue-specific read counts were generated for each AUGUSTUS gene model using the previously described TopHat alignments and the htseq-count command from HTSeq ("-m intersection-strict -a 3 -t exon -s no -i gene_id") (*47*). Reads were normalized for library size using DESeq v1.10.1 with default options (*48*). To facilitate examination of individual gene expression, library-normalized read counts from above were additionally normalized by transcript size to give "reads per kb transcript" (Table S1).

S. macrurus – Short reads from individual skeletal muscle and EO libraries were individually mapped to the Trinity transcriptome assembly using Bowtie (*49*) (v.0.12.8) with default parameters and read counting and ambiguity resolution were performed using RSEM (*50*) (v.1.2.3). Read counts were subsequently normalized for library size using the geometric mean method from DESeq (Table S1).

E. virescens, B. brachyistius – Expression data was calculated as for *S. macrurus* above (Table S1).

M. electricus – Short reads from skeletal muscle and EO libraries were independently mapped to the Trinity transcriptome assembly with RSEM (*50*), using the rsem-calculate-expression command, specifying paired-end data in addition to default parameters. All transcripts with mean read counts of < 10 across both EO and muscle were removed from analysis. Subsequently, read counts were normalized for library size using the geometric mean method from DEseq (Table S1) (*48*).

Sensitivity of E. electricus expression values

Rarefaction analysis: We performed a rarefaction analysis to evaluate the rate at which we could expect additional sequencing to result in detection of additional transcripts. To do this, we mapped all RNA-Seq reads from main EO in *E. electricus* to the SOAPdenovo2 genome assembly using TopHat (*42*) as previously described, and from here, sampled mapped reads representing 2%, 4%, 6%, 8%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of the total number of mapped reads in main EO, and generated read counts for all AUGUSTUS gene models using the htseq-count command from HTseq (*47*) as previously described. We looked at what number of genes met an

arbitrary threshold of 50 reads, 100 reads, and 1000 reads mapped to them in each of our 14 bins, and graphed our results (Figure S5). Our results indicated that the rate of new genes detected slows as a function of number of reads used, and our graph indicates that few new genes are detected with each incremental increase in read numbers as we approach 100% of reads used. Based on this, we feel confident that we have achieved sufficient sensitivity given available resources.

Incremental changes to read counts: In order to determine the effect of incremental changes in numbers of reads used on the stability of our gene expression values, we mapped all RNA-Seq reads from each of the *E. electricus* tissues independently to the SOAPdenovo2 genome assembly using TopHat (*42*) as previously described. Bins containing mapped reads were created by randomly sampling mapped read equally from each of the eight tissues, to reach final read counts of 2, 4, 6, 8, 10, 20, 30, 40, 60, 70, 80, 90, and 100 million reads. Read counts for each of our AUGUSTUS gene models were generated using the htseq-count command from HTseq as previously described, and we generated expression values in FPKM for each of our bins. We looked at what percentage of genes fell within 5%, 10%, 15%, 20%, 40%, 60%, and 80% of the expression values for the next highest bin (for example, expression values at 40 million reads were compared to that of 50 million; the 100 million read bin was compared to a second independently-sampled 100 million read bin). In order to reduce some of the noise from genes that were either not expressed, or very lowly expressed in our analysis, we removed first genes that had FPKM expression values < 0.05 in the two 100 million read bins (26,227 genes considered), and second, genes that had FPKM expression values of < 0.05 in any of the read bins (22,276 genes considered) (Figure S6). Our resulting graphs indicate that incremental changes in read depth have decreasing effects on the stability of expression values as you approach 100 million reads. In other words, incremental changes do not have a large effect on our gene expression values, indicating that our expression values are relatively stable.

8. K-mean Clustering of E. electricus Gene Expression

Genes with an average normalized read count across tissues of less than 10 were removed from further consideration. A variance stabilizing transformation was applied to the remaining genes using DESeq, followed by median centering. Low-information genes were removed using a local R implementation of the SUMCOV variance filter as described previously (*51*). The remaining approximately 6,000 genes were subjected to k-means clustering in R (*52*) using the stats package to discover tissue-specific expression signatures (Fig. S1). The value of k=12 was chosen after testing clustergrams, as well as iterative k-means clustering with a range of k values, to find a parameter which produced informative and stable clusters.

9. Assignment of Orthologs Between Electric Fish

S. macrurus - The one-to-one gene assignment table between *E. electricus* and *D. rerio* was used to identify orthologous genes from *S. macrurus*. *S. macrurus* transcripts assembled from Trinity were compared separately against all genes from *E. electricus* and *D. rerio* using BLAST (*44*) (blastn, e-value: 1E-5). Subsequently, all *S. macrurus* matches with each *E. electricus* gene were ranked from highest to lowest BLAST score. Another similar table was created for each *D. rerio* gene. If the same *S. macrurus* transcript matched a pair of genes from the *E. electricus* and *D. rerio* one-to-one assignment table, it was selected as the *S. macrurus* ortholog for the pair. If different transcripts matched a pair, they were both aligned with both *E. electricus* and *D. rerio* genes using CLUSTAL and the one with highest percentage match among four alignments was selected as the *S. macrurus* ortholog. Further manual confirmation of the matches was performed for the set of genes discussed in the manuscript by aligning translated amino acid or nucleotide sequences from the electric fishes and various other species (primarily zebrafish: *D. rerio*; channel catfish: *Ictalurus punctatus*; tilapia: *Oreochromis niloticus*; stickleback: *Gasterosteus aculeatus*; western clawed frog: *Xenopus tropicalis*; human: *Homo sapiens*), and generated gene trees using a maximum likelihood criterion using SeaView (*53*).

E. virescens - An identical procedure as *S. macrurus* was used to assign orthologous genes in *E. virescens*.

M. electricus- An identical procedure as *S. macrurus* was used to assign orthologous genes in *M. electricus*.

B. brachyistius - The *B. brachyistius* SM/EO paired-end assembly contained 147,923 assembled transcript contigs. Of these, 62,417 had homologues in the NCBI nr database using blastx and a cutoff of $1E^{-10}$. Of the 147,923 sequences for *B. brachyistius*, 49,902 had a homologous sequence in *E. electricus* (using tblastx, cutoff $1E^{-10}$). About 69% of the reduced set of *E. electricus* transcripts had a match in *B. brachyistius* sequences. Reciprocal blast between the reduced *E. electricus* transcriptome and the *B. brachyistius* transcriptome yields 7,960 matches between the two.

10. Analysis of Muscle Protein Regulation in Electric Organs

Data analysis approach

The transcripts of more than 30 muscle genes that are directly involved in muscle contraction, i.e., sarcomere, T-tubule, and sarcoplasmic reticulum (SR) (*54-56*), were identified in the transcriptomes of *E. electricus*, *S. macrurus*, *E. virescens*, *B. brachyistius*, and *M. electricus*. Gene identities for each transcript were derived using SeaView (*53*) (v.4.4.0) with ClustalW (*57*) alignments of the protein-coding regions of the electric fish transcripts and teleost orthologs retrieved from the Ensembl database

(*44*). The expression level for each gene was analyzed, and the abundance of each transcript in EO relative to that in skeletal muscle (SM) was computed for each species.

Results

Expression of most contraction-related genes was detected in all five species of electric fish and is shown in Table S2 and Figure S3. Transcripts with very low expression after normalization or those not detected included nebulette, JSRP1 (junctional sarcoplasmic reticulum protein 1), and HRC (histidine rich calcium binding protein), and are not shown. The relative transcript profiles of contraction-related genes were similar in all three EOs of *E. electricus*, i.e., main, Hunter's and Sachs' and showed that >65% of these genes were down-regulated at least 4-fold in the EOs compared to skeletal muscle. The transcript profiles found in *B. brachyistius* and *M. electricus* were similar to that of *E. electricus* in that most (~66% for both species) contraction-related genes were down-regulated in the EO by at least 4-fold. In contrast to both *E. electricus*, *B. brachyistius*, and *M. electricus*, the EO and skeletal muscle of *S. macrurus* showed similar expression levels in ~90% of contraction-related genes analyzed, whereas only one gene (troponin-C1) was found to be down-regulated by at least 4-fold in EO. The EO of *E. virescens* showed downregulation of contraction-related genes to a lesser extent than that found in *E. electricus*, *B. brachyistius*, and *M. electricus* but greater than in *S. macrurus*. Specifically, ~56% of these were down-regulated by at least 4-fold compared to skeletal muscle whereas ~42% showed similar expression. In the EOs of all five species less than 5% of contraction-related genes were found to be up-regulated by 4-fold or more compared to skeletal muscle. The transcript profiles of contraction-related genes found in the EOs of these five species are consistent with the expression profiles of the myogenic regulatory factor (MRF) genes of the MyoD family (Fig. S2), which regulate the transcription of many of these contraction-related genes in mammalian skeletal muscle since MRFs are strongly down-regulated in EOs of *E. electricus*, *B. brachyistius*, and *M. electricus* less so in the EO of *E. virescens*, and not down-regulated in the EO of *S. macrurus* (Fig. S2). These data indicate a transcriptional program that is more similar between the EO and skeletal muscle of *S. macrurus* than that observed in *E. electricus*, *B. brachyistius*, or *M. electricus*. These observations are intriguing in view of the evolutionary relationships between these electric fish species. Despite sharing a common gymnotiform ancestry, the contractile muscle gene expression profiles in EOs of *E. electricus* and *S. macrurus* differ considerably, whereas the EO of the distantly related *B. brachyistius* and *M. electricus* exhibits both contraction-related and MRF gene expression patterns that are similar to those in *E. electricus* but not those in *S. macrurus*. Further, these contraction-related gene expression patterns cannot be used to predict the cellular ultrastructure of electrocytes in these electric fish species. Previous studies have shown that fully differentiated electrocytes of *B. brachyistius* contain myofilamentous structures (*58, 59*) whereas mature electrocytes of *S. macrurus* do not (*5*, *60*, *61*). Myofilamentous structures are observed in electrocytes of young but not adult *E. electricus* (*62, 63*). In sum, these data indicate that transcriptional repression of the myogenic program is not a requisite for the emergence and maintenance of the EO phenotype.

11. Author Contributions

GAU, HZ, JRG, and JSA provided organisms; GAU, HZ, JRG, and LLT isolated the RNA and DNA; JRG and LLT designed and supervised the preparation involved in sequencing RNA and DNA; JDV, JRG, LLT, GNP, and MPS conceived and executed the data analysis pipeline; CDN, GAU, GBW, HM, HZ, JDV, JRG, LLT, MP, MPS, P-H C, RA, and RG analyzed the data; CDN, HM, JRG, LLT, MPS, and MRS designed the experiments; JRG, LLT, HZ and MRS wrote the manuscript and all authors edited the manuscript; MPS and MRS supervised this project.

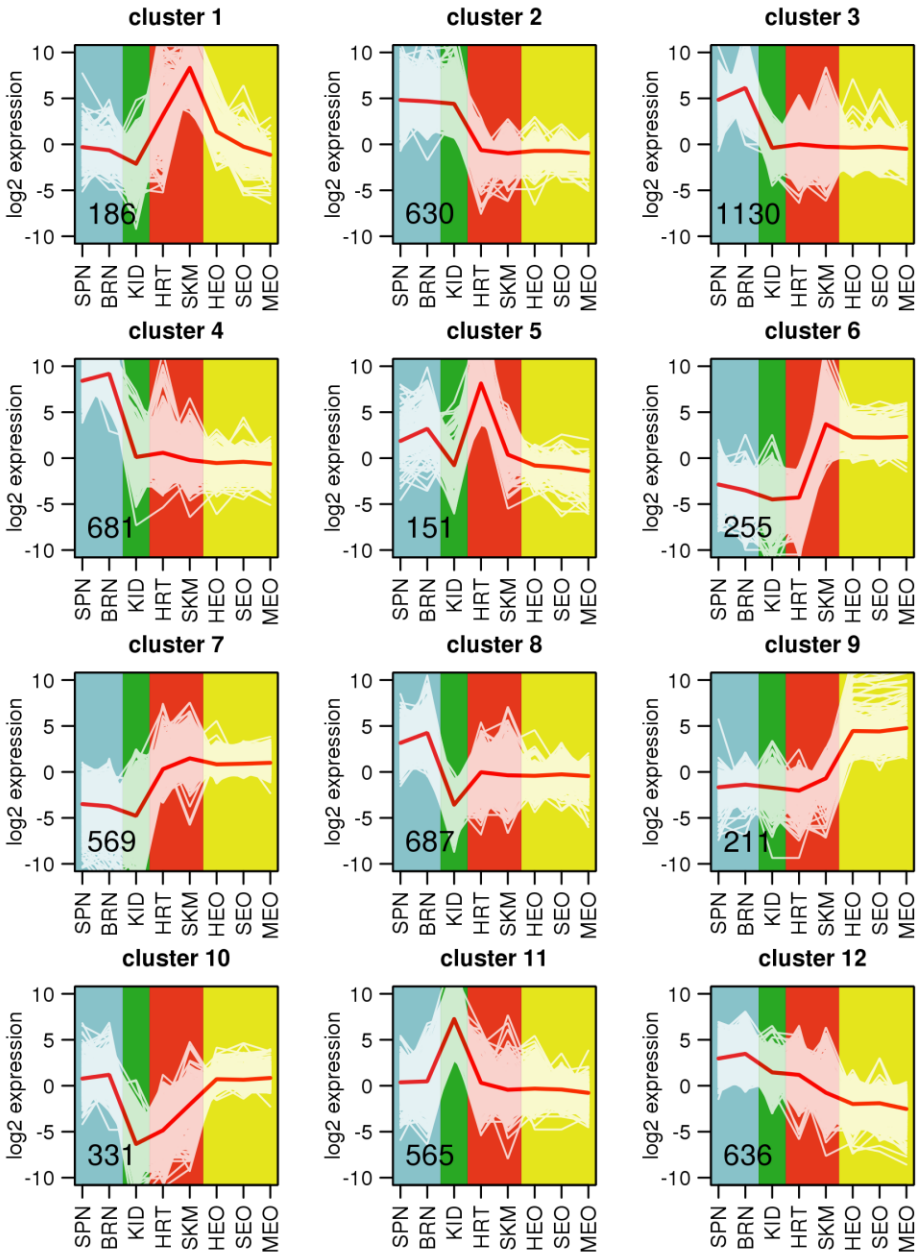


Fig. S1.

Clustering of co-expressed genes in *E. electricus*. A k-means clustering analysis (k=12) was performed as described in the Supplementary Text. Values in lower-left indicate the number of genes in each cluster. White plot lines represent individual genes and red plot lines show median values for the cluster. Background shading indicates general categories of tissue/cell type. SPN=spinal cord; BRN=brain; KID=kidney; HRT=heart; SKM=skeletal muscle; HEO=Hunter's EO; SEO=Sachs' EO; MEO=main EO.

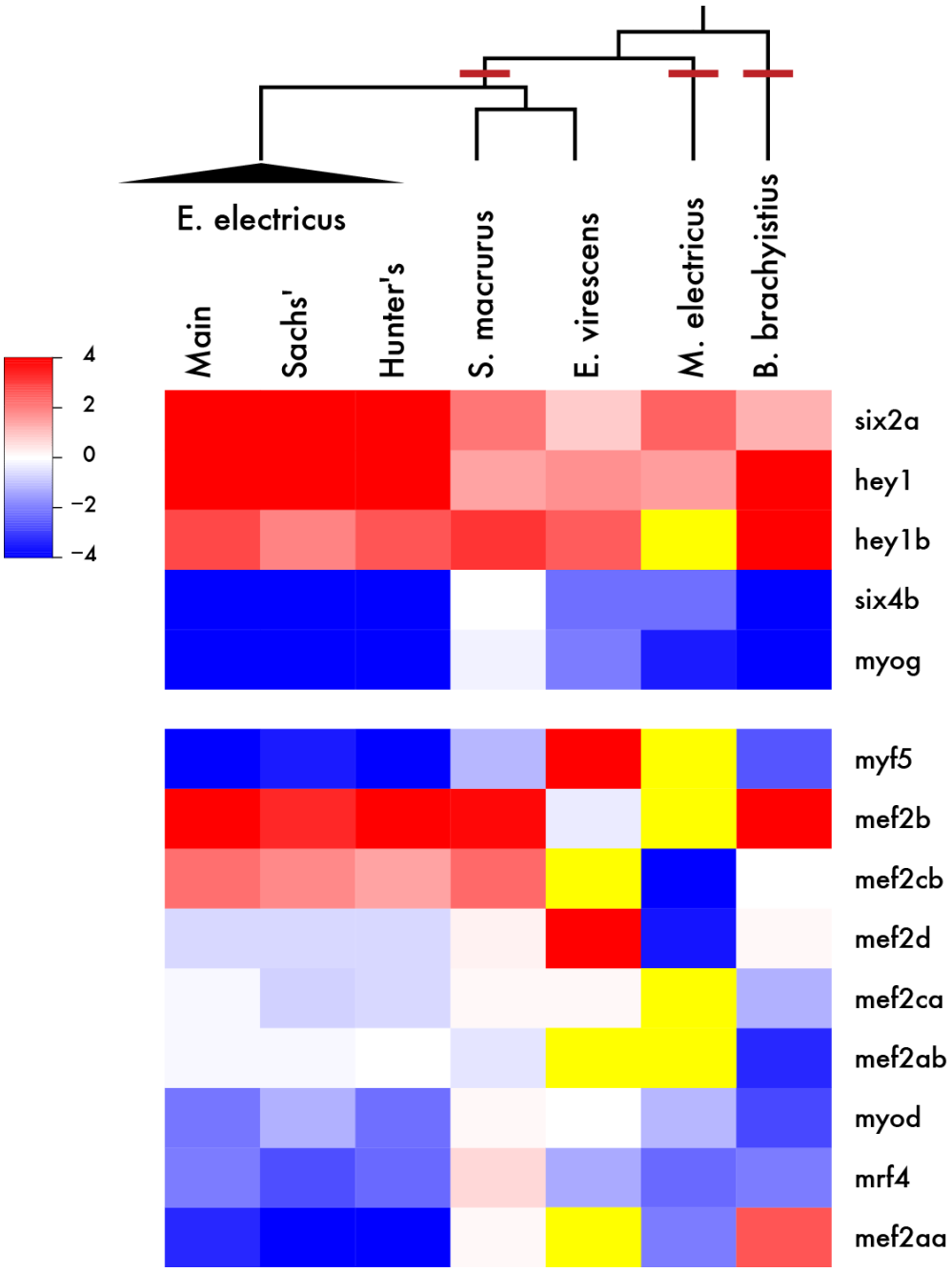


Fig. S2

Heatmap of myogenic transcription factor and related muscle development gene expression. Log2 transformed ratios of electric organ to skeletal muscle expression are displayed. Red colors indicate genes highly upregulated in electric organ, blue colors indicate genes highly downregulated in electric organ for each species. Yellow colors indicate the transcript was not detected. Note that myogenin is the sole "classical" transcription factor that is consistently down-regulated in all species.

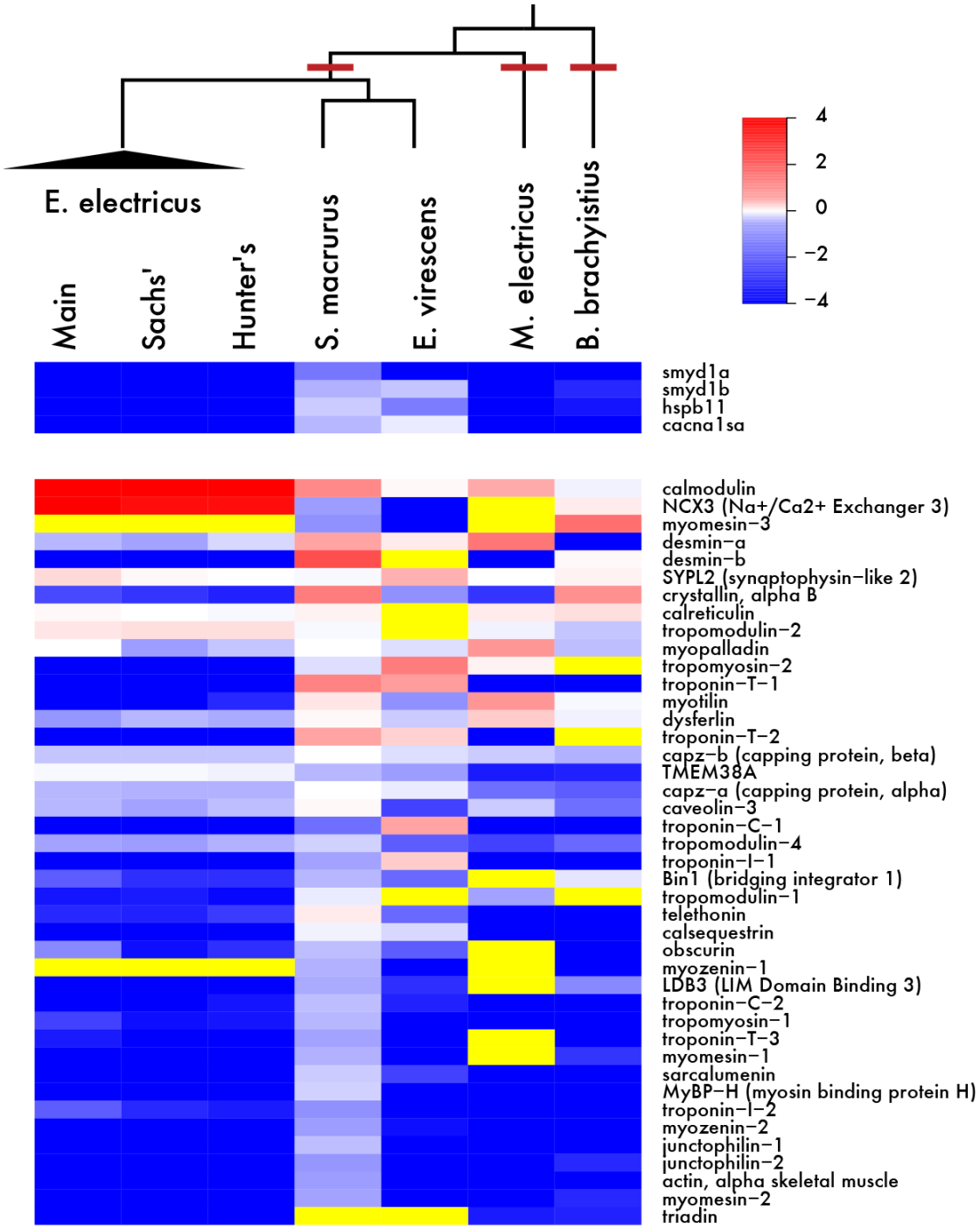


Fig. S3

Heatmap of sarcomeric gene expression. Log2 transformed ratios of electric organ to skeletal muscle expression are displayed. Red colors indicate genes highly upregulated in electric organ, blue colors indicate genes highly downregulated in electric organ for each species. Yellow indicates the transcript was not detected.

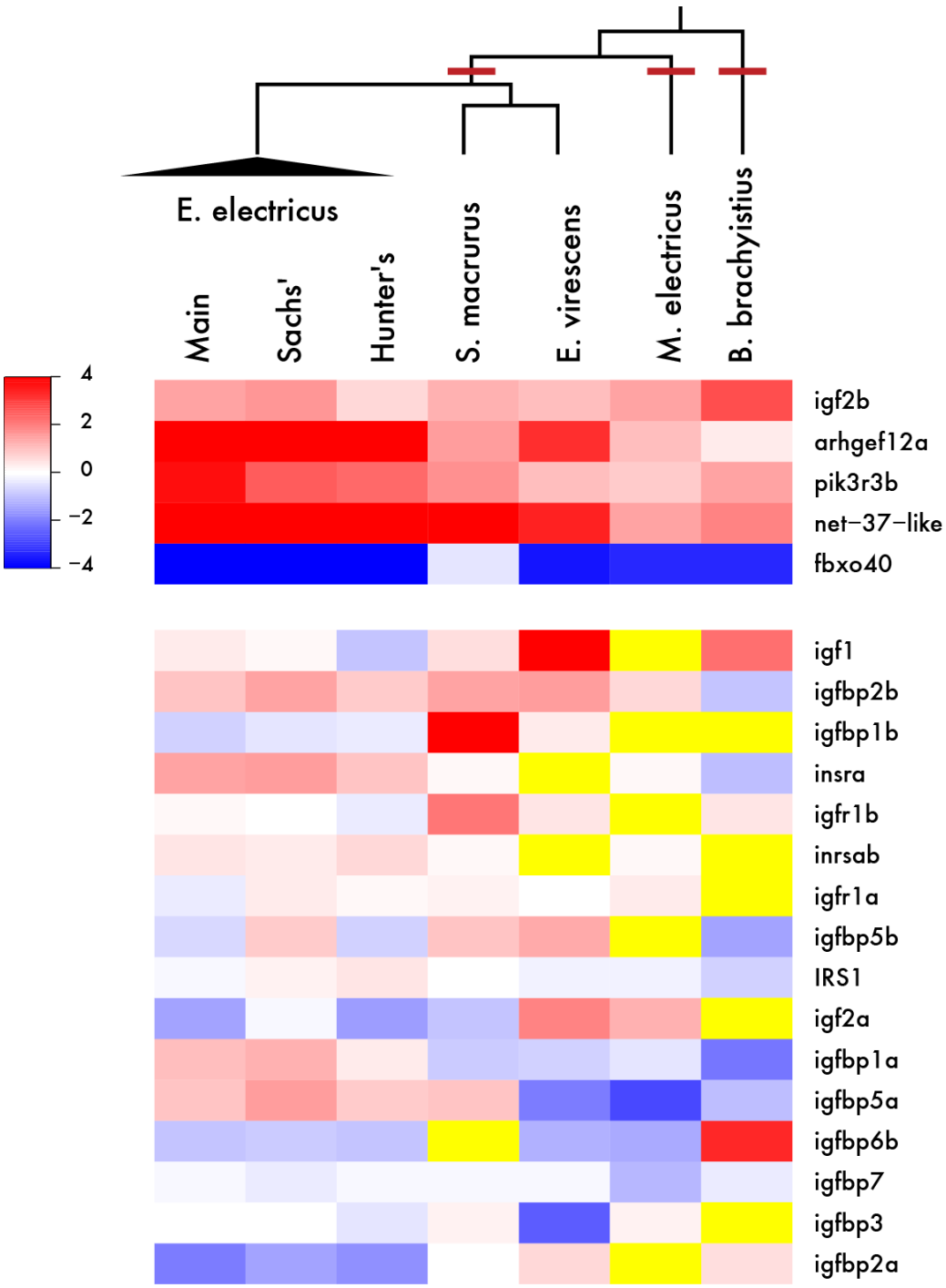


Fig. S4

Heatmap of IGF signaling gene expression. Log2 transformed ratios of electric organ to skeletal muscle expression are displayed. Red colors indicate genes highly upregulated in electric organ, blue colors indicate genes highly downregulated in electric organ for each species, yellow indicates the transcript was not detected. The set of genes that are similarly regulated in all species and that are illustrated in Fig. 2 are igf2b, arhgef21a, pik3r3b, NET37-like, and fbxo40.

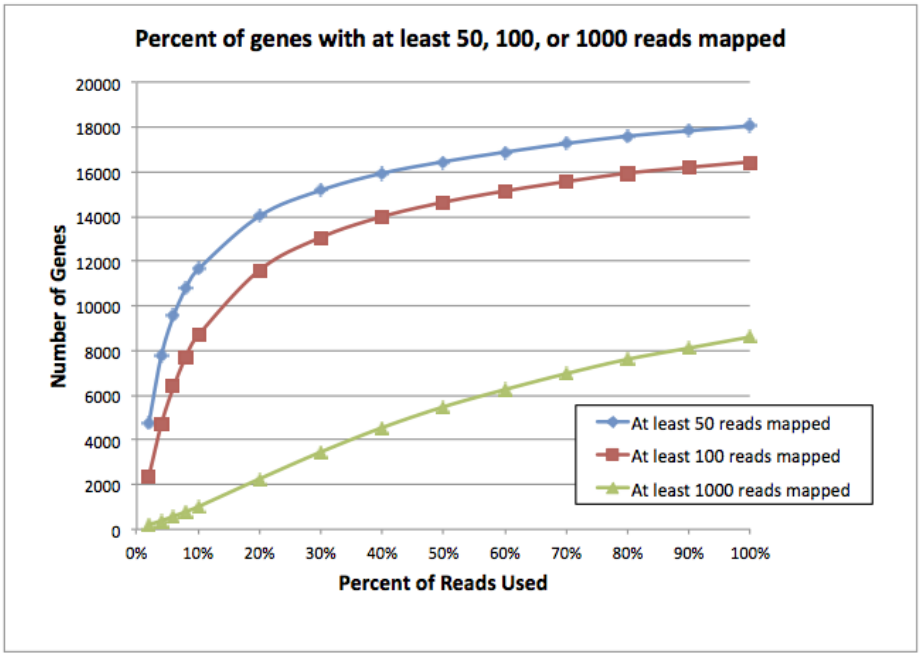
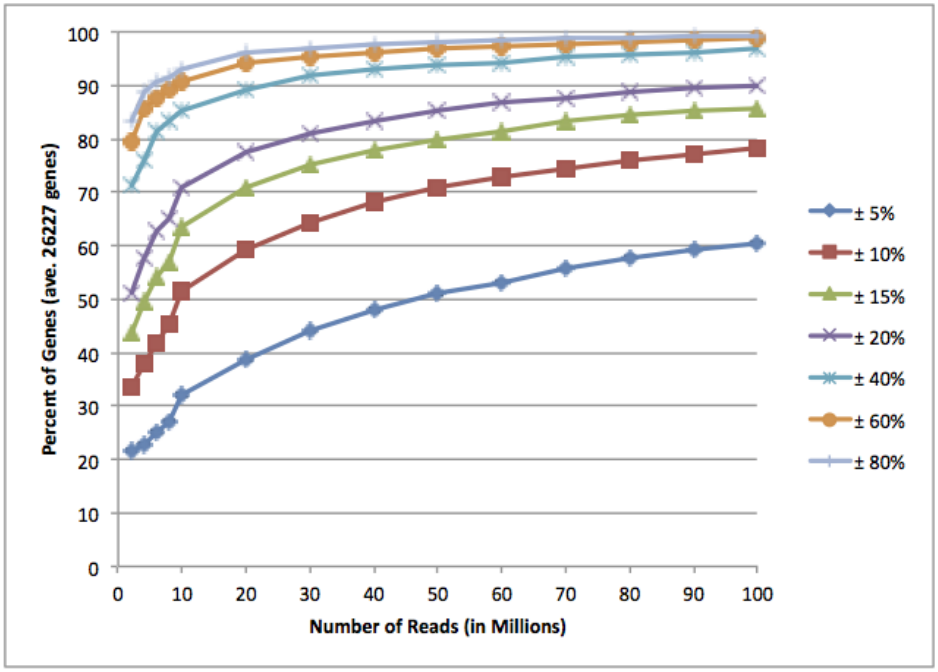


Fig. S5

Rarefaction analysis of additional sequencing depth on transcript content. We mapped all RNA-Seq reads from main EO in eel to the SOAPdenovo2 genome assembly using TopHat (*42*) as previously described, and from here, sampled mapped reads representing 2%, 4%, 6%, 8%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of the total number of mapped reads in main EO, and generated read counts for all AUGUSTUS gene models using the htseq-count command from HTseq (*43*) as previously described. We looked at what number of genes met an arbitrary threshold of 50 reads, 100 reads, and 1000 reads mapped to them in each of our 14 bins.

S6a:



S6b:

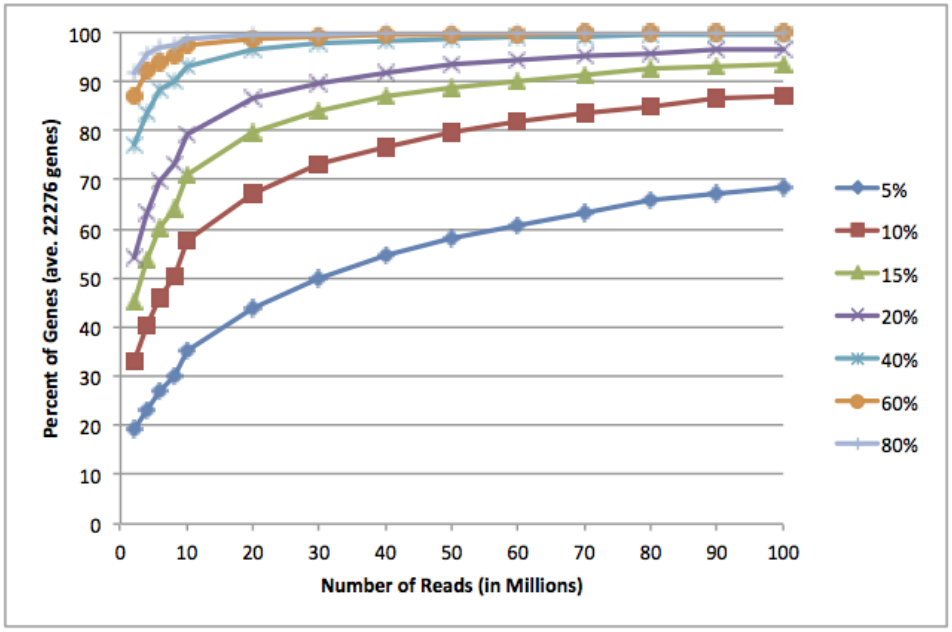


Fig. S6

Effect of sequencing depth on expression values Bins containing mapped reads were created by randomly sampling mapped reads equally from each of the eight tissues, to reach final read counts of 2, 4, 6, 8, 10, 20, 30, 40, 60, 70, 80, 90, and 100 million reads. Read counts for each of our AUGUSTUS gene models were generated using the htseq-count command from HTseq as previously described, and we generated expression values in FPKM for each of our bins.

Table S1 (separate file)

Gene expression in electric fishes. Combined table of gene expression the eight tissues (brain, spinal cord, heart, skeletal muscle, kidney, main EO, Sachs' EO, Hunter's EO) of *E. electricus*, (b) Gene expression in skeletal muscle and EO of *S. macrurus*, (c) Gene expression in skeletal muscle and EO of *E. virescens*, (d) Gene expression in skeletal muscle and EO samples of *B. brachyistius*.

Table S2 (separate file)

Relative expression abundance of muscle genes in electric organ relative to skeletal muscle in four species of electric fish. Colors indicate overexpression (red) and underexpression (green) in electric organ relative to skeletal muscle. n-d indicates transcripts not found or ratios not determined because abundance was too low for accurate quantitation.

Table S3.

Summary of literature regarding transcription factor interactions in figure 2b. For each gene, gene name, whether it is expressed in presomatic mesoderm/somites is listed, along with citation, whether it is expressed in mature muscle along with citation, whether expression is known to inhibit muscle development along with citation, and various notes about expression patterns and interactions are listed.

Gene Name	Expressed in PSM/Somite	Expressed in mature muscle	Expression inhibits muscle development	Notes
hey1a	yes (*64*)	no (*64*)	yes (*10*, *65*)	Hey1 is a corepressor of Myod (*66*), myogenin and mef2c (*10*). Myogenin and myoD activate actin (*67*), and myoD activates myosin (*68*). Myogenin activates smyd1 (*69*)
hey1b	yes (*64*)	no (*64*)	yes (*10*, *66*)	
six2a	yes (*70*)			ARE motif in Na+/K+ ATPase bound by Six/sine oculus family genes (*8*, *71*)
six1a	yes (*72*, *73*)	yes-mice (*72*)	no--needed in fast fibers (*73*, *74*)	ARE motif in Na+/K+ ATPase bound by Six/sine oculus family genes (*8*, *71*)
six4b	yes (*75*)	no (*75*)	no--needed in fast fibers (*73*, *74*)	ARE motif in Na+/K+ ATPase bound by Six/sine oculus family genes (*8*, *71*)

Table S4.

Summary of literature regarding protein interactions and locations in figure 2c. For each protein (where known), its location in skeletal muscle, known binding partners, results of knockout experiments in other vertebrates, and roles in myopathy are listed. Notes summarize other potentially relevant features of these proteins.

Gene Name	Location	Binding Partner	Phenotype of KO/Mutation	Myopathy	Notes
smyd1	M line (*14*, *76*)	myosin (*14*, *76*)	disrupts sarcomere (*14*, *76*, *77*)		Smyd proteins stabilize sarcomeric proteins (*76-79*). Smyd1 is activated by myogenin (*79*, *80*)
smyd2b	I band (*78*, *79*)	titin (*78*)	disrupts sarcomere (*78*)		Smyd proteins stabilize sarcomeric proteins (*76-79*)
hspb11			disrupts Sarcomere (*14*)		
fbxo40		IRSI (*81*)	enlarged myofibers (*81*)		Fbxo40 complex limits IGF1 signaling by inducing IRIS1 ubiquination (*81*). The activated IRIS1/IGFR complex inhibits protein degradation by inhibiting Foxo (*21*)
IGF2b	secreted by cell in autocrine fashion (*19*, *80*)	IGF1 receptor (*80*)	muscle atrophy (*80*) severe hypoplasia (*82*)	Attenuated IGF signaling associated with cardiac myopathy (*83*).	Enhances body size and developmental rate in zebrafish (*84*). Overexpression prevents muscle atrophy in mice (*85*). IGF activates IRIS1, IRIS1/IGFR complex binds to PIK3 (*20*). This activated complex can influence cell proliferation and increase protein synthesis by inhibiting Foxo (*21*)

Table S4 (continued)

ARHGEF12	cytoplasm (*86*)	IGF1 receptor (*86*)			
pik3r3b	cytoplasm (*87*)	IRS1 (*87*)			
NET37-like	nuclear envelope (*23*)	IGF2 (*23*)	prevents muscle differentiation, decreases myogenin (*23*)		Increases autocrine release of IGFII (*23*)
col6a6	connective tissue, basal lamina (*88*)	col6a3 (*88*)		Muscular Dystrophy (*88*)	
col14a1	connective tissue, basal lamina (*89*)	type I collagen (*89*)	prevents glycosylation of dystroglycan (*89*)		
dmd	plasma membrane (*90*)	dystroglycan (*90*)		muscular dystrophy (*90*)	
glytl1b	Golgi apparatus (*91*)	alpha dystroglycan (*91*)		muscular dystrophy (*91*)	

Table S5.

Summary of genomic sequencing experiments. Genomic DNA from *E. electricus* was sequenced using Illumina technologies.

Platform	library type	median read length (nt)	number of reads (millions)
Illumina GAIIx	paired-end	73	37
Illumina HiSeq 2000	paired-end	99	600
Illumina HiSeq 2000	2k mate-pair	51	830

Table S6.
Summary of transcriptomic sequencing experiments. RNA-Seq experiments were performed in *E. electricus, S. macrurus, E. virescens, M. electricus* and *B. brachyistius*.

Species	samples	sequencing technology	run type
E. electricus	3 tissues - main EO, Sachs' EO, skeletal muscle	Illumina GAIIx	paired end
E. electricus	7 tissue - brain, spinal cord, main EO, Sachs' EO, Hunter's EO, skeletal muscle, heart	Illumina HiSeq 2000	single end
E. electricus	1 tissue - kidney	Illumina HiSeq 2000	paired end
S. macrurus	2 tissues - EO and skeletal muscle	Illumina HiSeq 2000	paired end
E. virescens	2 tissues - EO and skeletal muscle	Illumina HiSeq 2000	paired end
E. virescens	2 tissues - EO and skeletal muscle	Illumina HiSeq 2000	paired end
B. brachyistius	2 tissues - EO and skeletal muscle	Illumina HiSeq 2000	paired end
M. electricus	2 tissues - EO and skeletal muscle	Illumina HiSeq 2000	paired end

Table S7.
Summary of *E. electricus* genome assembly. Summary statistics were calculated for (a) both gapped (containing Ns) and ungapped (Ns stripped) builds of SOAPdenovo2 assembly.

	Gapped	Ungapped
Number of Contigs	121323	121323
Total Length	560202223	523576209
Average length	4617.4	4315.6
Longest Contig	996512	972603
Shortest Contig	100	100
Contig N50	104253	103026
Contig N90	12699	15371
G/C Content	42.50%	42.50%

References and Notes

1. J. S. Albert, W. G. R. Crampton, “Electroreception and electrogenesis.” in *The Physiology of Fishes* (Springer, New York, ed. 3, 2005), pp. 431–472.

2. C. Darwin, in *On the Origin of Species by Means of Natural Selection* (J. Murray, London, 1859), pp. ix, 1.

3. B. A. Block, Evolutionary novelties: How fish have built a heater out of muscle. *Am. Zool.* **31**, 726–742 (1991).

4. H. Kokubo, S. Tomita-Miyagawa, Y. Hamada, Y. Saga, Hesr1 and Hesr2 regulate atrioventricular boundary formation in the developing heart through the repression of Tbx2. *Development* **134**, 747–755 (2007). Medline doi:10.1242/dev.02777

5. G. A. Unguez, H. H. Zakon, Reexpression of myogenic proteins in mature electric organ after removal of neural input. *J. Neurosci.* **18**, 9924–9935 (1998). Medline

6. Materials and methods are available as supplementary materials on *Science* Online.

7. H. Ghanbari, H. C. Seo, A. Fjose, A. W. Brändli, Molecular cloning and embryonic expression of *Xenopus Six* homeobox genes. *Mech. Dev.* **101**, 271–277 (2001). Medline doi:10.1016/S0925-4773(00)00572-4

8. F. Spitz, J. Demignon, A. Porteu, A. Kahn, J. P. Concordet, D. Daegelen, P. Maire, Expression of myogenin during embryogenesis is controlled by Six/*sine oculis* homeoproteins through a conserved MEF3 binding site. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14220–14225 (1998). Medline doi:10.1073/pnas.95.24.14220

9. K. Kawakami, S. Noguchi, M. Noda, H. Takahashi, T. Ohta, M. Kawamura, H. Nojima, K. Nagano, T. Hirose, S. Inayama, H. Hayashida, T. Miyata, S. Numa, Primary structure of the α-subunit of *Torpedo californica* ($Na^+ + K^+$)ATPase deduced from cDNA sequence. *Nature* **316**, 733–736 (1985). Medline doi:10.1038/316733a0

10. M. F. Buas, S. Kabak, T. Kadesch, The Notch effector Hey1 associates with myogenic target genes to repress myogenesis. *J. Biol. Chem.* **285**, 1249–1258 (2010). Medline doi:10.1074/jbc.M109.046441

11. K. J. Nowak, K. E. Davies, Duchenne muscular dystrophy and dystrophin: Pathogenesis and opportunities for treatment. *EMBO Rep.* **5**, 872–876 (2004). Medline doi:10.1038/sj.embor.7400221

12. R. G. Berry, S. Despa, W. Fuller, D. M. Bers, M. J. Shattock, Differential distribution and regulation of mouse cardiac Na^+/K^+-ATPase α_1 and α_2 subunits in T-tubule and surface sarcolemmal membranes. *Cardiovasc. Res.* **73**, 92–100 (2007). Medline doi:10.1016/j.cardiores.2006.11.006

13. J. Lowe, G. M. Araujo, A. R. Pedrenho, N. Nunes-Tavares, M. G. Ribeiro, A. Hassón-Voloch, Polarized distribution of Na^+, K^+-ATPase α-subunit isoforms in electrocyte membranes. *Biochim. Biophys. Acta* **1661**, 40–46 (2004). Medline doi:10.1016/j.bbamem.2003.11.020

14. N. Klüver, L. Yang, W. Busch, K. Scheffler, P. Renner, U. Strähle, S. Scholz, Transcriptional response of zebrafish embryos exposed to neurotoxic compounds reveals a muscle

activity dependent *hspb11* expression. *PLOS ONE* **6**, e29063 (2011). Medline doi:10.1371/journal.pone.0029063

15. C. Duan, H. Ren, S. Gao, Insulin-like growth factors (IGFs), IGF receptors, and IGF-binding proteins: Roles in skeletal muscle growth and differentiation. *Gen. Comp. Endocrinol.* **167**, 344–351 (2010). Medline doi:10.1016/j.ygcen.2010.04.009

16. B. C. Hoopes, M. Rimbault, D. Liebers, E. A. Ostrander, N. B. Sutter, The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm. Genome* **23**, 780–790 (2012). Medline doi:10.1007/s00335-012-9417-z

17. N. B. Sutter, C. D. Bustamante, K. Chase, M. M. Gray, K. Zhao, L. Zhu, B. Padhukasahasram, E. Karlins, S. Davis, P. G. Jones, P. Quignon, G. S. Johnson, H. G. Parker, N. Fretwell, D. S. Mosher, D. F. Lawler, E. Satyaraj, M. Nordborg, K. G. Lark, R. K. Wayne, E. A. Ostrander, A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112–115 (2007). Medline doi:10.1126/science.1137045

18. D. J. Emlen, I. A. Warren, A. Johns, I. Dworkin, L. C. Lavine, A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons. *Science* **337**, 860–864 (2012). Medline doi:10.1126/science.1224286

19. A. S. Van Laere, M. Nguyen, M. Braunschweig, C. Nezer, C. Collette, L. Moreau, A. L. Archibald, C. S. Haley, N. Buys, M. Tally, G. Andersson, M. Georges, L. Andersson, A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836 (2003). Medline doi:10.1038/nature02064

20. I. Mothe, L. Delahaye, C. Filloux, S. Pons, M. F. White, E. Van Obberghen, Interaction of wild type and dominant-negative p55PIK regulatory subunit of phosphatidylinositol 3-kinase with insulin-like growth factor-1 signaling proteins. *Mol. Endocrinol.* **11**, 1911–1923 (1997). Medline doi:10.1210/mend.11.13.0029

21. A. Otto, K. Patel, Signalling and the control of skeletal muscle size. *Exp. Cell Res.* **316**, 3059–3066 (2010). Medline doi:10.1016/j.yexcr.2010.04.009

22. I. H. Chen, M. Huber, T. Guan, A. Bubeck, L. Gerace, Nuclear envelope transmembrane proteins (NETs) that are up-regulated during myogenesis. *BMC Cell Biol.* **7**, 38 (2006). Medline doi:10.1186/1471-2121-7-38

23. K. Datta, T. Guan, L. Gerace, NET37, a nuclear envelope transmembrane protein with glycosidase homology, is involved in myoblast differentiation. *J. Biol. Chem.* **284**, 29666–29676 (2009). Medline doi:10.1074/jbc.M109.034041

24. S. E. Mate, K. J. Brown, E. P. Hoffman, Integrated genomics and proteomics of the *Torpedo californica* electric organ: Concordance with the mammalian neuromuscular junction. *Skeletal Muscle* **1**, 20 (2011). Medline doi:10.1186/2044-5040-1-20

25. M. E. Alfaro, F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, L. J. Harmon, Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13410–13414 (2009). Medline doi:10.1073/pnas.0811087106

26. C. Linneaus, *Systema Naturae* (Holmiae, ed. XII, 1766).

27. J. S. Albert, *Check List of the Freshwater Fishes of South and Central America* (Edipuers, Porto Alegre, Brazil, 2003).

28. A. Valenciennes, “Les poissons.” in *Le Règne Animal Distribué d'Après son Organisation, pour Servir de Base à l'Historie Naturelle des Animaux, et d'Introduction a l'Anatomie Comparée*, G. Cuvier, Ed. (Masson, Paris, ed. 3, 1838–1842).

29. K. R. Bradnam, J. N. Fass, A. Alexandrov, P. Baranay, M. Bechner, I. Birol, S. Boisvert, J. A. Chapman, G. Chapuis, R. Chikhi, H. Chitsaz, W.-C. Chou, J. Corbeil, C. Del Fabbro, T. R. Docking, R. Durbin, D. Earl, S. Emrich, P. Fedotov, N. A. Fonseca, G. Ganapathy, R. A. Gibbs, S. Gnerre, É. Godzaridis, S. Goldstein, M. Haimel, G. Hall, D. Haussler, J. B. Hiatt, I. Y. Ho, J. Howard, M. Hunt, S. D. Jackman, D. B. Jaffe, E. Jarvis, H. Jiang, S. Kazakov, P. J. Kersey, J. O. Kitzman, J. R. Knight, S. Koren, T.-W. Lam, D. Lavenier, F. Laviolette, Y. Li, Z. Li, B. Liu, Y. Liu, R. Luo, I. MacCallum, M. D. MacManes, N. Maillet, S. Melnikov, B. M. Vieira, D. Naquin, Z. Ning, T. D. Otto, B. Paten, O. S. Paulo, A. M. Phillippy, F. Pina-Martins, M. Place, D. Przybylski, X. Qin, C. Qu, F. J. Ribeiro, S. Richards, D. S. Rokhsar, J. G. Ruby, S. Scalabrin, M. C. Schatz, D. C. Schwartz, A. Sergushichev, T. Sharpe, T. I. Shaw, J. Shendure, Y. Shi, J. T. Simpson, H. Song, F. Tsarev, F. Vezzi, R. Vicedomini, J. Wang, K. C. Worley, S. Yin, S.-M. Yiu, J. Yuan, G. Zhang, H. Zhang, S. Zhou, I. F. Korf, Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. http://arxiv.org/abs/1301.5406 (2013).

30. J. F. Gmelin, *Systema Naturae per Regna Tria Naturae, Secundum Classes, Ordines, Genera, Species; cum Characteribus, Differentiis, Synonymis, Locis*. Editio decimo tertia, aucta, reformata. 3 vols. in 9 parts. (Lipsiae, 1788–1993), vol. 1 (pt 3), pp. 1033–1516.

31. T. Gregory, Animal Genome Size Database (2013); www.genomesize.com.

32. D. C. Hardie, P. D. N. Hebert, The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* **46**, 683–706 (2003). Medline doi:10.1139/g03-040

33. P. Chomczynski, N. Sacchi, The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: Twenty-something years on. *Nat. Protoc.* **1**, 581–585 (2006). Medline doi:10.1038/nprot.2006.83

34. D. C. Hardie, P. D. N. Hebert, Genome-size evolution in fishes. *Can. J. Fish. Aquat. Sci.* **61**, 1636–1646 (2004). doi:10.1139/f04-106

35. R. Hinegardner, Evolution of cellular DNA content in teleost fishes. *Am. Nat.* **102**, 517–523 (1968). doi:10.1086/282564

36. R. Hinegardner, D. E. Rosen, Cellular DNA content and the evolution of teleostean fishes. *Am. Nat.* **106**, 621–644 (1972). doi:10.1086/282801

37. J. T. Simpson, Exploring genome characteristics and sequence quality without a reference. http://arxiv.org/abs/1307.8026 (2013).

38. A. E. Vinogradov, Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. *Cytometry* **31**, 100–109 (1998). Medline doi:10.1002/(SICI)1097-0320(19980201)31:2<100::AID-CYTO5>3.0.CO;2-Q

39. R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S. M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T. W. Lam, J. Wang, SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012). Medline doi:10.1186/2047-217X-1-18

40. T. D. Wu, C. K. Watanabe, GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005). Medline doi:10.1093/bioinformatics/bti310

41. G. Parra, K. Bradnam, I. Korf, CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007). Medline doi:10.1093/bioinformatics/btm071

42. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009). Medline doi:10.1093/bioinformatics/btp120

43. M. Stanke, B. Morgenstern, AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33** (suppl. 2), W465–W467 (2005). Medline doi:10.1093/nar/gki458

44. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). Medline doi:10.1016/S0022-2836(05)80360-2

45. P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. K. Kähäri, D. Keefe, S. Keenan, R. Kinsella, M. Komorowska, G. Koscielny, E. Kulesha, P. Larsson, I. Longden, W. McLaren, M. Muffato, B. Overduin, M. Pignatelli, B. Pritchard, H. S. Riat, G. R. Ritchie, M. Ruffier, M. Schuster, D. Sobral, Y. A. Tang, K. Taylor, S. Trevanion, J. Vandrovcova, S. White, M. Wilson, S. P. Wilder, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernández-Suarez, J. Harrow, J. Herrero, T. J. Hubbard, A. Parker, G. Proctor, G. Spudich, J. Vogel, A. Yates, A. Zadissa, S. M. Searle, Ensembl 2012. *Nucleic Acids Res.* **40**, D84–D90 (2012). Medline doi:10.1093/nar/gkr991

46. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). Medline doi:10.1038/nbt.1883

47. S. Anders, P. T. Pyl, W. Huber, HTSeq – A Python framework to work with high-throughput sequencing data. http://biorxiv.org/content/early/2014/02/20/002824 (2014); doi: 10.1101/002824.

48. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010). Medline doi:10.1186/gb-2010-11-10-r106

49. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). Medline doi:10.1186/gb-2009-10-3-r25

50. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). Medline doi:10.1186/1471-2105-12-323

51. D. Tritchler, E. Parkhomenko, J. Beyene, Filtering genes for cluster and network analysis. *BMC Bioinformatics* **10**, 193 (2009). Medline doi:10.1186/1471-2105-10-193

52. R Development Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2008); www.R-project.org.

53. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010). Medline doi:10.1093/molbev/msp259

54. L. Al-Qusairi, J. Laporte, T-tubule biogenesis and triad formation in skeletal muscle and implication in human diseases. *Skeletal Muscle* **1**, 26 (2011). Medline doi:10.1186/2044-5040-1-26

55. S. Lange, E. Ehler, M. Gautel, From A to Z and back? Multicompartment proteins in the sarcomere. *Trends Cell Biol.* **16**, 11–18 (2006). Medline doi:10.1016/j.tcb.2005.11.007

56. S. Treves, M. Vukcevic, M. Maj, R. Thurnheer, B. Mosca, F. Zorzato, Minor sarcoplasmic reticulum membrane components that modulate excitation-contraction coupling in striated muscles. *J. Physiol.* **587**, 3071–3079 (2009). Medline doi:10.1113/jphysiol.2009.171876

57. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007). Medline doi:10.1093/bioinformatics/btm404

58. A. H. Bass, in *Electroreception* (Wiley, New York, 1986), pp. 13–70.

59. J. R. Gallant, C. D. Hopkins, D. L. Deitcher, Differential expression of genes and proteins between electric organ and skeletal muscle in the mormyrid electric fish *Brienomyrus brachyistius*. *J. Exp. Biol.* **215**, 2479–2494 (2012). Medline doi:10.1242/jeb.063222

60. I. R. Schwartz, G. D. Pappas, M. V. Bennett, The fine structure of electrocytes in weakly electric teleosts. *J. Neurocytol.* **4**, 87–114 (1975). Medline doi:10.1007/BF01099098

61. G. A. Unguez, H. H. Zakon, Phenotypic conversion of distinct muscle fiber populations to electrocytes in a weakly electric fish. *J. Comp. Neurol.* **399**, 20–34 (1998). Medline doi:10.1002/(SICI)1096-9861(19980914)399:1<20::AID-CNE2>3.0.CO;2-C

62. M. A. Esquibel, I. Alonso, H. Meyer, C. Chagas, G. de Castro, [Some aspects of the histogenesis and ontogenesis of electric organs in *Electrophorus electricus* (L.)]. *C. R. Acad. Sci. Hebd. Seances Acad. Sci. D* **273**, 196–199 (1971). Medline

63. R. D. Machado, W. de Souza, G. Cotta-Pereira, G. de Oliveira Castro, On the fine structure of the electrocyte of *Electrophorus electricus* L. *Cell Tissue Res.* **174**, 355–366 (1976). Medline doi:10.1007/BF00220681

64. C. Winkler, H. Elmasri, B. Klamt, J. N. Volff, M. Gessler, Characterization of *hey* bHLH genes in teleost fish. *Dev. Genes Evol.* **213**, 541–553 (2003). Medline doi:10.1007/s00427-003-0360-6

65. L. Sun, H. Xie, M. A. Mori, R. Alexander, B. Yuan, S. M. Hattangadi, Q. Liu, C. R. Kahn, H. F. Lodish, *Mir193b-365* is essential for brown fat differentiation. *Nat. Cell Biol.* **13**, 958–965 (2011). Medline doi:10.1038/ncb2286

66. J. Sun, C. N. Kamei, M. D. Layne, M. K. Jain, J. K. Liao, M. E. Lee, M. T. Chin, Regulation of myogenic terminal differentiation by the hairy-related transcription factor CHF2. *J. Biol. Chem.* **276**, 18591–18596 (2001). Medline doi:10.1074/jbc.M101163200

67. G. E. Muscat, J. Emery, E. S. Collie, Tissue-specific expression of the skeletal alpha-actin gene involves sequences that can function independently of MyoD and Id. *Gene Expr.* **2**, 241–257 (1992). Medline

68. J. D. Meissner, P. K. Umeda, K. C. Chang, G. Gros, R. J. Scheibe, Activation of the beta myosin heavy chain promoter by MEF-2D, MyoD, p300, and the calcineurin/NFATc1 pathway. *J. Cell. Physiol.* **211**, 138–148 (2007). Medline doi:10.1002/jcp.20916

69. D. Li, Z. Niu, W. Yu, Y. Qian, Q. Wang, Q. Li, Z. Yi, J. Luo, X. Wu, Y. Wang, R. J. Schwartz, M. Liu, SMYD1, the myogenic activator, is a direct target of serum response factor and myogenin. *Nucleic Acids Res.* **37**, 7059–7071 (2009). Medline doi:10.1093/nar/gkp773

70. G. Oliver, A. Mailhos, R. Wehr, N. G. Copeland, N. A. Jenkins, P. Gruss, *Six3*, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development* **121**, 4045–4055 (1995). Medline

71. K. Kawakami, H. Ohto, K. Ikeda, R. G. Roeder, Structure, function and expression of a murine homeobox protein AREC3, a homologue of *Drosophila sine oculis* gene product, and implication in development. *Nucleic Acids Res.* **24**, 303–310 (1996). Medline doi:10.1093/nar/24.2.303

72. B. Thisse, C. Thisse, Fast release clones: A high throughput expression analysis. *ZFIN Direct Data Submission* (2004); http://zfin.org.

73. D. A. Bessarab, S. W. Chong, B. P. Srinivas, V. Korzh, Six1a is required for the onset of fast muscle differentiation in zebrafish. *Dev. Biol.* **323**, 216–228 (2008). Medline doi:10.1016/j.ydbio.2008.08.015

74. A. F. Richard, J. Demignon, I. Sakakibara, J. Pujol, M. Favier, L. Strochlic, F. Le Grand, N. Sgarioto, A. Guernec, A. Schmitt, N. Cagnard, R. Huang, C. Legay, I. Guillet-Deniau, P. Maire, Genesis of muscle fiber-type diversity during mouse embryogenesis relies on Six1 and Six4 gene expression. *Dev. Biol.* **359**, 303–320 (2011). Medline doi:10.1016/j.ydbio.2011.08.010

75. M. Kobayashi, H. Osanai, K. Kawakami, M. Yamamoto, Expression of three zebrafish Six4 genes in the cranial sensory placodes and the developing somites. *Mech. Dev.* **98**, 151–155 (2000). Medline doi:10.1016/S0925-4773(00)00451-2

76. S. Just, B. Meder, I. M. Berger, C. Etard, N. Trano, E. Patzel, D. Hassel, S. Marquart, T. Dahme, B. Vogel, M. C. Fishman, H. A. Katus, U. Strähle, W. Rottbauer, The myosin-interacting protein SMYD1 is essential for sarcomere organization. *J. Cell Sci.* **124**, 3127–3136 (2011). Medline doi:10.1242/jcs.084772

77. X. Tan, J. Rotllant, H. Li, P. De Deyne, S. J. Du, SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2713–2718 (2006). Medline doi:10.1073/pnas.0509503103

78. L. T. Donlin, C. Andresen, S. Just, E. Rudensky, C. T. Pappas, M. Kruger, E. Y. Jacobs, A. Unger, A. Zieseniss, M. W. Dobenecker, T. Voelkel, B. T. Chait, C. C. Gregorio, W. Rottbauer, A. Tarakhovsky, W. A. Linke, Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. *Genes Dev.* **26**, 114–119 (2012). Medline doi:10.1101/gad.177758.111

79. H. Li, Y. Zhong, Z. Wang, J. Gao, J. Xu, W. Chu, J. Zhang, S. Fang, S. J. Du, Smyd1b is required for skeletal and cardiac muscle function in zebrafish. *Mol. Biol. Cell* **24**, 3511–3521 (2013). Medline doi:10.1091/mbc.E13-06-0352

80. S. Schiaffino, C. Mammucari, Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: Insights from genetic models. *Skeletal Muscle* **1**, 4 (2011). Medline doi:10.1186/2044-5040-1-4

81. J. Shi, L. Luo, J. Eash, C. Ibebunjo, D. J. Glass, The SCF-Fbxo40 complex induces IRS1 ubiquitination in skeletal muscle, limiting IGF1 signaling. *Dev. Cell* **21**, 835–847 (2011). Medline doi:10.1016/j.devcel.2011.09.011

82. J. P. Liu, J. Baker, A. S. Perkins, E. J. Robertson, A. Efstratiadis, Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell* **75**, 59–72 (1993). Medline

83. L. Elia, R. Contu, M. Quintavalle, F. Varrone, C. Chimenti, M. A. Russo, V. Cimino, L. De Marinis, A. Frustaci, D. Catalucci, G. Condorelli, Reciprocal regulation of microRNA-1 and insulin-like growth factor-1 signal transduction cascade in cardiac and skeletal muscle in physiological and pathological conditions. *Circulation* **120**, 2377–2385 (2009). Medline doi:10.1161/CIRCULATIONAHA.109.879429

84. X. Wang, L. Lu, Y. Li, M. Li, C. Chen, Q. Feng, C. Zhang, C. Duan, Molecular and functional characterization of two distinct IGF binding protein-6 genes in zebrafish. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **296**, R1348–R1357 (2009). Medline doi:10.1152/ajpregu.90969.2008

85. E. R. Barton-Davis, D. I. Shoturma, A. Musaro, N. Rosenthal, H. L. Sweeney, Viral mediated expression of insulin-like growth factor I blocks the aging-related loss of skeletal muscle function. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15603–15607 (1998). Medline doi:10.1073/pnas.95.26.15603

86. S. Taya, N. Inagaki, H. Sengiku, H. Makino, A. Iwamatsu, I. Urakawa, K. Nagao, S. Kataoka, K. Kaibuchi, Direct interaction of insulin-like growth factor-1 receptor with leukemia-associated RhoGEF. *J. Cell Biol.* **155**, 809–820 (2001). Medline doi:10.1083/jcb.200106139

87. P. R. Shepherd, Mechanisms regulating phosphoinositide 3-kinase signalling in insulin-sensitive tissues. *Acta Physiol. Scand.* **183**, 3–12 (2005). Medline doi:10.1111/j.1365-201X.2004.01382.x

88. W. R. Telfer, A. S. Busta, C. G. Bonnemann, E. L. Feldman, J. J. Dowling, Zebrafish models of collagen VI-related myopathies. *Hum. Mol. Genet.* **19**, 2433–2444 (2010). Medline doi:10.1093/hmg/ddq126

89. H. L. Bader, E. Lambert, A. Guiraud, M. Malbouyres, W. Driever, M. Koch, F. Ruggiero, Zebrafish collagen XIV is transiently expressed in epithelia and is required for proper function of certain basement membranes. *J. Biol. Chem.* **288**, 6777–6787 (2013). Medline doi:10.1074/jbc.M112.430637

90. J. Berger, P. D. Currie, Zebrafish models flex their muscles to shed light on muscular dystrophies. *Dis. Model. Mech.* **5**, 726–732 (2012). Medline doi:10.1242/dmm.010082

91. M. Brockington, S. Torelli, P. Prandini, C. Boito, N. F. Dolatshad, C. Longman, S. C. Brown, F. Muntoni, Localization and functional analysis of the LARGE family of glycosyltransferases: Significance for muscular dystrophy. *Hum. Mol. Genet.* **14**, 657–665 (2005). Medline doi:10.1093/hmg/ddi062