Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/7/e1700523/DC1

Supplementary Materials for

A tail of two voltages: Proteomic comparison of the three electric organs of the electric eel

Lindsay L. Traeger, Grzegorz Sabat, Gregory A. Barrett-Wilt, Gregg B. Wells, Michael R. Sussman

Published 5 July 2017, *Sci. Adv.* **3**, e1700523 (2017)
DOI: 10.1126/sciadv.1700523

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at advances.sciencemag.org/cgi/content/full/3/7/e1700523/DC1)

- table S7 (Microsoft Excel format). Phosphopeptides in electric organ discharge–related proteins that differ in abundance compared to protein abundance.
- table S8 (Microsoft Excel format). Raw output from Proteome Discoverer, unenriched, whole-proteome samples.
- table S9 (Microsoft Excel format). Raw output from Proteome Discoverer, titanium dioxide–enriched phosphopeptides.
- table S10 (Microsoft Excel format). Correlation of RNA and protein abundance values.
- table S12 (Microsoft Excel format). Median-normalized channel intensity values for titanium dioxide–enriched phosphopeptides.

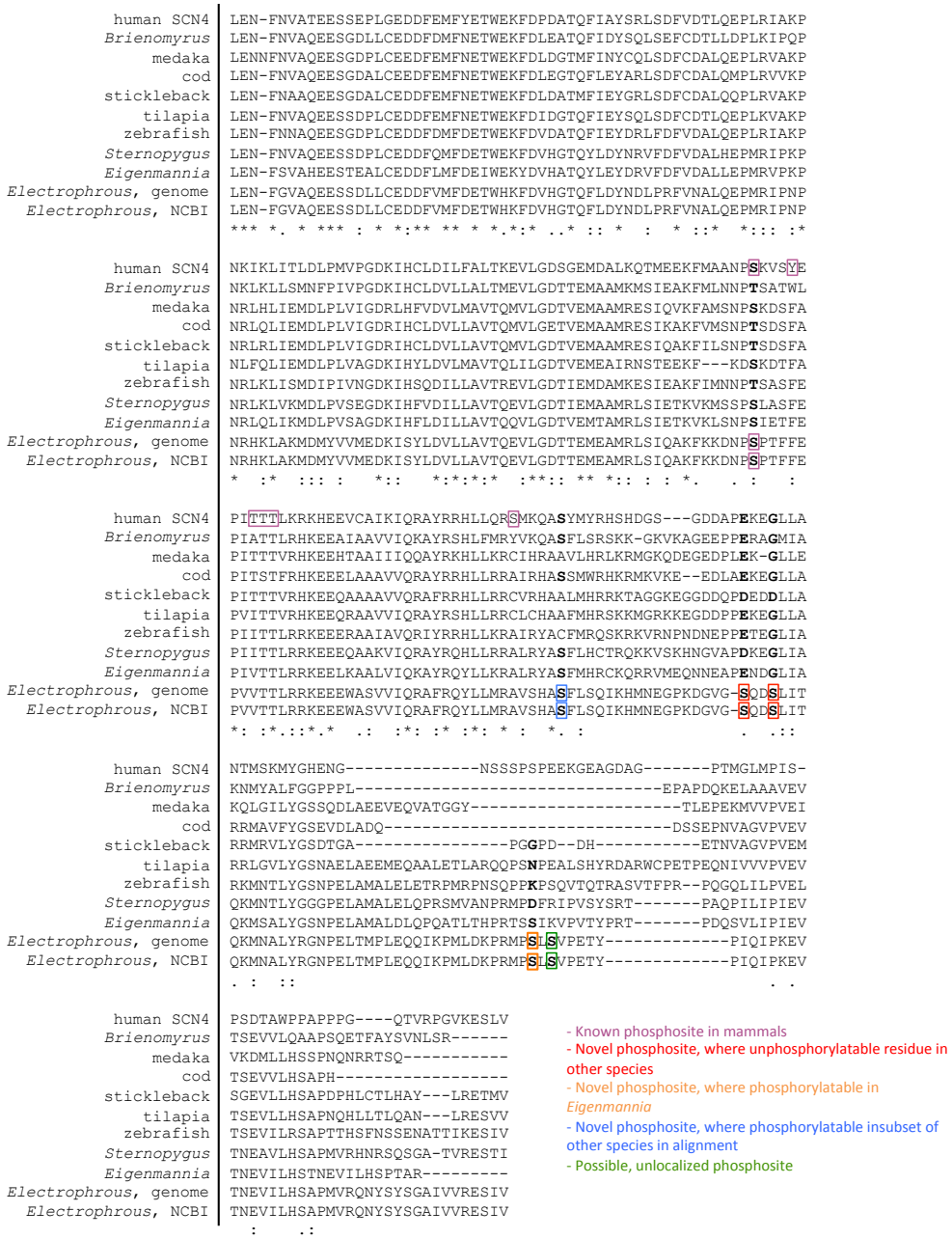
human SCN4
Brienomyrus
medaka
cod
stickleback
tilapia
zebrafish
Sternopygus
Eigenmannia
Electrophrous, genome
Electrophrous, NCBI
- Known phosphosite in mammals
- Novel phosphosite, where unphosphorylatable residue in other species
- Novel phosphosite, where phosphorylatable in Eigenmannia
- Novel phosphosite, where phosphorylatable insubset of other species in alignment
- Possible, unlocalized phosphosite

fig. S1. Phosphorylation sites in the C-terminal domain of *E. electricus* SCN4aa. C-terminal domain of the protein shows variation among species, including among Gymnotiformes (*Eigenmannia virescens* and *Sternopygus macrurus*). The SCN4aa sequence available at NCBI for *E. electricus* was included for reference. One phosphosite previously described in mammals and four novel phosphosites were localized in the C-terminus of SCN4aa, and are depicted in colored boxes. Red boxes indicated the two localized phosphosites at residues in which other species in the alignment have unphosphorylatable residues. The blue box indicates a localized phosphorylation site at a residue that is also phosphorylatable in a subset of the species included in the alignment (phosphorylatable in the Gymnotiformes the Mormyroid (*Brienomyrus brachyistius*) included in the alignment). The orange box indicates a localized phosphosite at a residue that is phosphorylatable in *E. virescens* (Human shows a phosphorylated residue, however, the alignment is very poor for human). The green box indicates an unlocalized phosphosite possibility. Sequences downloaded from NCBI: Human SCN4a, as ancestral reference (NP_000325.4) and *Sternopygus macrurus* (AAK55442.2). Sequences downloaded from ENSEMBL Build 81: tilapia (ENSONIP00000009933), cod (ENSGMOP00000003144), medaka (ENSORLP00000014568), stickleback (ENSGACP00000004617), and zebrafish (ENSDARP00000134593). Sequences from transcriptome assemblies (*6*) for and transcript sequences for *Brienomyrus brachyistius* (transcript comp27150_c0_seq1) and *Eigenmannia virescens* (transcript Ev-comp269953_c0_seq20). Transcript assemblies covering a significant portion of the C-terminal end were not identified in the transcriptome assemblies for *Malapterurus electricus* or *B. brachyistius*.

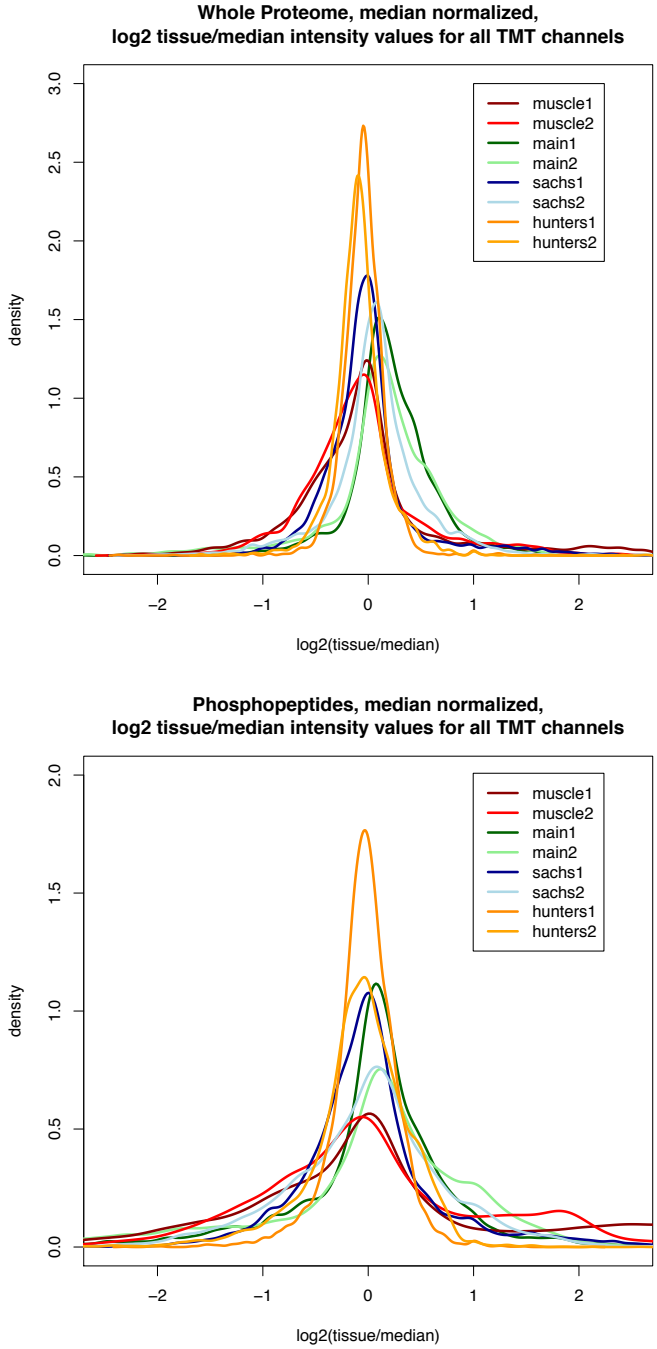

fig. S2. Normalization of proteomic and phosphoproteomic data. Density plots showing distribution of $\log_2$ (tissue/median) normalized intensity values.

table S1. Novel and known phosphosites in *E. electricus* proteins. Contains phosphosite information for a subset of proteins discussed in the manuscript. A phosphosite was considered localized if it had a localization score of 75% or greater. Whether a phosphosite was considered known in mammals or novel was determined based on protein alignments with *E. electricus*, human, mouse, and zebrafish sequences. A phosphosite in *E. electricus* was considered novel if it was at least five amino acids away from a known phosphosite (this information is recorded in the "notes" column).

table S2. Median-normalized channel intensity values for unenriched, whole-proteome samples.

table S3. Normalized intensity ratios, $\log_2$ (tissue/median), on a per-peptide basis.

table S4. Potential roles for nonelectrogenic proteins and phosphopeptides differentially abundant in electric organs. Contains list of proteins that were identified as differentially abundant in main or Sachs' electric organ protein or phosphopeptide clusters.

table S5. Expression values (RNA) for all predicted genes in assembly. Expression values are in "reads per kilobase transcript", as described in the methods. Raw reads from (*6*), and include brain, spinal cord, whole heart, skeletal muscle, main electric organ, Sachs' electric organ, Hunter's electric organ, and whole kidney.

table S6. Intensity ratios, $\log_2$ (tissue/median), on a per-protein group basis.

table S7. Phosphopeptides in electric organ discharge–related proteins that differ in abundance compared to protein abundance. Shown in this table are the log2 (tissue/median)

values of each differentially abundant phosphopeptide in all eight tissues. For these phosphopeptides the relative abundance of the phosphopeptide and protein are shown, as well as the differences in relative abundance between phosphopeptides and proteins. Abundance differences were considered significant if the difference between the phosphopeptide abundance and the protein abundance was at least two fold (difference in $\log_2$ values of at least 1). Orange highlighted rows in "note" column indicate other phosphopeptides that differ in abundance, but have missed cleavages, and so were not discussed further.

table S8. Raw output from Proteome Discoverer, unenriched, whole-proteome samples. Values are unnormalized raw channel intensities, for input into custom script for normalization and quantitation.

table S9. Raw output from Proteome Discoverer, titanium dioxide–enriched phosphopeptides.

table S10. Correlation of RNA and protein abundance values. All abundance values for RNA or protein are in log2 (electric organ/muscle). "Group" values in each table have following meanings: 1: not differentially expressed (DE) in RNA or protein, 2: DE in protein only, 3: DE in RNA only, 4: DE in both RNA and protein, and 5: DE in both RNA and protein, but in opposite directions. (A) Eel1 correlation values, RNA expression values and protein abundance values, for all three electric organs. (B) Eel 2 correlation values, RNA expression values, and protein abundance values, for all three electric organs. (C) main electric organ, the join between tabs 1 and 2, where in both biological replicates, the gene models showed the same pattern (D) Join for Sachs' electric organ, (E) Join for Hunter's electric organ.

table S11. Comparison of new genome assembly and gene annotations to the previous assembly. (A) Table shows statistics on comparing various assemblies to the original assembly published previously (*6*). After incorporating new 454 reads, and a round of scaffolding and gapfilling, this improved assembly has ~3.5 times fewer contigs, 2.5 times fewer scaffolds, the longest contig is ~7 times longer, the longest scaffold is 3.5 times longer. In other words, more of the newest assembly is in longer assembled pieces. Statistics generated by Quast (*55*) for each genome assembly rendition (see methods): "SOAP genome only" (original genome assembly (*6*)), "scaffolded SOAP genome, illumina only" (original genome assembly scaffolded with Illumina mate-pair reads only), "gapfilled, scaffolded SOAP genome, illumina only" (scaffolded with illumina plus a round of gap filling), "scaffolded, with illumina and 454" (original genome assembly scaffolded with both Illumina and 454 mate-pair reads), "gapfilled, scaffolded with illumina and 454" (scaffolded with illumina/454 plus a round of gap filling. This is the assembly used in this manuscript). (**B)** Comparison of gene models generated in old and new assemblies. Predicted proteins were blasted against zebrafish proteins, and compared to one another based on blast hit store. This comparison revealed that ~84% of the time, the new predicted protein sequences were as good or better than the previous gene model, meaning 84% of the time, they received the same blast hit score to zebrafish, or a better score compared to the previous equivalent gene model.

Table S11

A

SOAP genome only

Assembly	contigs (split on Ns)	scaffolds
# contigs (>= 0 bp)	471054	121323
# contigs (>= 1000 bp)	103972	16715
Largest contig	65235	886512
N50	5954	107900

scaffolded SOAP genome, illumina only

Assembly	contigs (split on Ns)	scaffolds
# contigs (>= 0 bp)	353998	8840
# contigs (>= 1000 bp)	102528	6133
Largest contig	65235	3119439
N50	6126	595548

gapfilled, scaffolded SOAP genome, illumina only

Assembly	contigs (split on Ns)	scaffolds
# contigs (>= 0 bp)	47656	8840
# contigs (>= 1000 bp)	32881	6096
Largest contig	429623	3115075
N50	37312	596078

scaffolded, with illumina and 454

Assembly	contigs (split on Ns)	scaffolds
# contigs (>= 0 bp)	353965	8788
# contigs (>= 1000 bp)	102523	6081
Largest contig	65235	3119439
N50	6127	616204

gapfilled, scaffolded with illumina and 454

Assembly	contigs	scaffolds
# contigs (>= 0 bp)	47607	8788
# contigs (>= 1000 bp)	32849	6044
Largest contig	438728	3115075
N50	37346	613956

B

Gene Model Comparison	Number of Proteins
Old and new gene models 'tie'	6714
New models have better score	7365
Old models have better score	2619

table S12. Median-normalized channel intensity values for titanium dioxide enriched phosphopeptides.

table S13. Peptide and protein group counts for unenriched, whole-proteome experiment, and TiOX-enriched phosphoproteome experiment.

Unenriched, whole proteome		
	# unique peptides	26668
	# protein groups	2873
TiOX enriched, phosphoproteome		
	# unique peptides	7905
	# unique phosphopeptides	4334
	% enrichment	54.80%
	# unique proteins	2076