INTRODUCTION
The electric eel (Electrophorus electricus) is a freshwater fish native to South America, most famous for its ability to produce high-voltage electric discharges that are used for predation and defense. The evolutionary history of electrogenic organs is particularly interesting, because electric organs have evolved not only in multiple independent fish lineages from myogenic precursors (1, 2); even within each lineage, there is a tremendous amount of variation in their function. This is particularly true in Gymnotiformes, where, for example, E. electricus is the only known species that has evolved three distinct electric organs and is capable of generating strong-voltage discharges (3). Like all other Gymnotiformes, E. electricus generates weak electric organ discharges from Sachs’ electric organ and the posterior portion of Hunter’s electric organ for the purposes of navigation and communication in their native habitat of murky fresh water. However, in addition to this function, E. electricus produces enormous electric organ discharges of tens to hundreds, up to 600 V and 1 A of current (4) from main electric organ and the anterior portion of Hunter’s electric organ for the purposes of predation, defense, and “remote controlling” prey—forcing hiding prey to reveal their location by using high-voltage discharges to induce muscle contractions (5). Because of this unusual ability to produce very high electric discharges, E. electricus and other strong-voltage species, such as Torpedo californica, are also rich sources for proteins involved in electric organ discharges.

Recently, we used DNA and RNA sequencing technologies and a comparative genomic approach to obtain a draft genome sequence of the first electric fish, the strong-voltage (E. electricus) eel. We hypothesized that these two electrogenic outputs have different energetic demands reflected by differences in their proteome and phosphoproteome. We report the use of isotope-assisted quantitative mass spectrometry to test this hypothesis. We observed novel phosphorylation sites in sodium transporters and identified a potassium channel with unique differences in protein concentration among the electric organs. In addition, we found transcription factors and protein kinases that show differential abundance in the strong versus weak electric organs. Our findings support the hypothesis that proteomic differences among electric organs underlie differences in energetic needs, reflecting a trade-off between generating weak voltages continuously and strong voltages intermittently.

RESULTS
Novel phosphorylation sites in essential sodium transport proteins involved in electrogenesis
Given the unique importance of electrogenic proteins (that is, sodium and potassium transporters) in electrocytes (Fig. 1A), we hypothesized that there may be phosphorylated residues not used for regulation in other animals (9) and that the uniquely high abundance of sodium channels and pumps in this tissue, together with the application of new MS methods for quantification, could reveal phosphosites that were not identified in studies with nerves and muscle in other organisms, including humans. We hypothesize that these sites not only are important for differences between the two types of electrogenic tissue observed in this fish but also may represent previously unknown means of regulating sodium fluxes involved in human behavioral and motor diseases.

1Department of Genetics, University of Wisconsin-Madison, Madison, WI 53706, USA.
2Biotechnology Center, University of Wisconsin-Madison, Madison, WI 53706, USA.
3Department of Molecular and Cellular Medicine, Texas A&M University, College Station, TX 77843, USA.
4Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA.
5Present address: Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, USA.
6†Corresponding author. Email: msussman@wisc.edu

Fig. 1. Locations of phosphorylated residues that were detected in representative members of four important electrogenic protein families. (A) Illustration of electrogenic proteins in electric organ discharge. Action potential is driven by massive Na$^+$ influx mediated by Na$^+$ channels (AChRs and voltage-gated Na$^+$ channels) in the innervated membrane face, whereas the Na$^+$/K$^+$-ATPase works to maintain the negative membrane potentials at the noninnervated face. This results in a large transcellular potential difference, and because electrocytes are arranged massively in series, like batteries in a flashlight, voltages summate. ADP, adenosine 5'-diphosphate; K$_{ir}$, inwardly rectifying potassium channel; Na$_v$, voltage-gated sodium channel. (B) Relative location of identified phosphorylated residues in electrogenic proteins. Green lines indicate the location of phosphorylated residues identified in these data that have been described in mammals (9). Red lines indicate novel phosphorylated residues in _E. electricus_. Blue indicates phosphosites identified in _E. electricus_ that fall within less than five residues of a known phosphosite in mammals. Amino acid (aa) length given is for gene model chosen for analysis (see table S1 for gene model selection, when more than one was possible, if applicable). (C) Phosphorylated residues (yellow or pink) of Na$^+$/K$^+$-ATPase ATP1a2a from _E. electrophorus_ are shown in their homologous positions in the crystal structure of the α subunit of Na$^+$/K$^+$-ATPase from _Sus scrofa_ (PDB accession 3WGU). Ribbon colors: red, nucleotide-binding domain; purple, phosphorylation domain; blue, actuator domain; green, membrane domain. (D) Phosphorylated residues (red or cyan) of voltage-gated Na$^+$ channel SCN4aa are shown in a transmembrane topology cartoon that is predicted for this _E. electrophorus_ protein (UniProt accession P02719), in lieu of a high-resolution three-dimensional structure that is not yet available. (E) Phosphorylated residues (red or cyan) of the nicotinic AChR β subunit CHRNB1 from _E. electrophorus_ are shown in a transmembrane topology cartoon that is predicted for this _E. electrophorus_ protein based on its homology with mouse CHRNB1 (UniProt accession P09690), also in lieu of a high-resolution three-dimensional structure. The phosphorylated residues are found intracellularly between transmembrane domains M3 and M4, a region whose structure is currently undefined in high-resolution structures of nicotinic AChRs. (F) Phosphorylated residues (yellow) of ACHE from _E. electrophorus_ are shown in their homologous positions in the crystal structure of ACHE from _T. californica_ (PDB accession 36FM). Color code in (C) to (F) for phosphorylated residues: yellow or red, phosphorylation at these residues is not found in mammals and is potentially relevant to unique activities of electrocytes; pink or cyan, phosphorylation at these residues is found in mammals.
in regulating sodium affinity in mammalian ATP1a1, and hypothesized to do the same in other isoforms based on homology (10), although, to the best of our knowledge, this phosphosite has never been directly observed in ATP1a2. This is likely because this region of the sodium/potassium adenosine triphosphatase (ATPase) is highly conserved, and enzymatic degradation with trypsin for subsequent MS analyses creates identical tryptic peptides for all commonly studied mammalian isoforms (ATP1a1, ATP1a2, and ATP1a3 all yield the peptide RNSVFQQGMK in humans, mice, and rats); thus, the resulting peptide sequence cannot be assigned to one ATPase isoform over another. The homologous region of E. electricus ATPases shows more variation, although it is still highly conserved, and the sequence in this region for ATP1a2 is distinct from all other E. electricus isoforms [RNS(p)VFQQGMR]. The voltage-gated sodium channel denoted as SCN4aa displayed 13 novel and 1 known phosphosite, most of which are cytoplasmically located between domains 1 and 2 and after S6 of domain 4 (Fig. 1D), whereas the sodium channel denoted as SCN4ab had five novel phosphosites, which are cytoplasmically located before domain 1, between domains 1 and 2, between domains 2 and 3, and after S6 of domain 4. Three nicotinic acetylcholine receptor (AChR) subunits displayed the following phosphosites, all of which were located cytoplasmically in the large loop between transmembrane domains 3 and 4: CHRNB1 had one novel and one known phosphosite (Fig. 1E); CHRNND had two novel and two known phosphosites, one of which is next to a phosphosite only reported in an electric fish species (T. californica) (9) that independently evolved electric organs from E. electricus, and CHRNG had one novel and one known phosphosite also only previously reported for T. californica. The acetylcholinesterase (ACHE) enzyme had two novel phosphosites separated by 20 residues but predicted to be close together in three-dimensional space (Fig. 1F). We were unable to detect any phosphosites for voltage-gated $K^+$ channels.

Calmodulin is highly abundant in the E. electricus electric organ, representing 2% of the total protein content of the electrolyte (11). Although it was among the most abundant proteins in main electric organ and strong-voltage section of Hunter’s electric organ, this was not the case in Sachs’ electric organ or muscle (table S2), suggesting that it plays an important role in the intermittent strong-voltage discharge. Despite its abundance, a clear role for calmodulin in electric organ function has not been described. One clue to its function may come from the demonstration that SCN4a, a voltage-gated Na$^+$ channel, is regulated by Ca$^{2+}$ and calmodulin through the interactions of calmodulin with the C-terminal domain of SCN4a (12). Further, the crystal structure for the ternary complex of a C-terminal domain of a voltage-gated Na$^+$ channel, calmodulin, and fibroblast growth factor 13 (FGF13) also highlights a role for FGF13 in voltage-gated Na$^+$ channel regulation via the C-terminal domain (13). Ten phosphopeptides matching different locations within FGF13 were identified in this study (table S3), highlighting a possible regulatory mechanism for the abundant voltage-gated Na$^+$ channel in the electric organ via the C terminus of the protein. Mutations in the C-terminal domain of SCN4a have implications in human disease (13).

Teleost fishes have duplicated SCN4a sodium channel genes, called SCN4aa and SCN4ab. In E. electricus, other Gymnotiformes, and mor- myroids (an independent lineage of electric fish), SCN4aa is highly expressed in the electric organ and has experienced a higher rate of molecular evolution compared to SCN4ab, which is highly expressed in both the muscle and electric organ. These features suggest that this Na$^+$ channel duplication was important in the independent evolution of electric organs (14). Using protein alignments among E. electricus, other Gymnotiformes, the mormyroid Brieniomyrus brachyistius, other teleost species, and humans, we observed that the C-terminal domain of SCN4aa is highly variable among species (fig. S1). Within the C-terminal domain of SCN4aa, we identified a total of five phosphosites in the E. electricus protein. Whereas one of these has been previously described in mammals (9), four are novel. Two novel phosphosites were at nonphosphorylatable residues in all other species in the alignment, which may indicate unique functions in E. electricus. The third novel phosphosite was localized to a residue that is phosphorylatable in a subset of other species, including all electric fish in the alignment (Eigennannia virescens, Sternopygus macrurus, and B. brachyistius), which may represent an important function in electric fishes. The fourth novel phosphosite in E. electricus is at a phosphorylatable residue in the protein alignment only in E. virescens and is a phosphomimetic D in S. macrurus. These novel phosphorylation targets might be functional adaptations in E. electricus and other electric fishes. The hypothesis that SCN4aa is co-regulated by interactions with calmodulin and FGF13 in electric organ remains to be experimentally tested.

**Unique patterns of potassium channel protein abundance across electric organs**

Diversity in the abundance and types of ion channels present in electrocytes has been suggested to be a driver for electrogenic signal diversity in Gymnotiformes (15) and, by extension, speciation. Therefore, it follows that because E. electricus has three distinct electric organs that serve distinct electrogenic functions, perhaps the same underlying principles apply within a species as they do between species. Here, we first examined proteins known to be important for propagating the electric organ discharge and identified several ion channel and ion pump proteins that followed a general pattern of being most abundant in main electric organ, least abundant in Sachs’ electric organ, and of moderate abundance for Hunter’s electric organ including several Na$^+/K^+$-ATPase subunits (for example, ATP1a1a), voltage-gated Na$^+$ channel subunits (for example, SCN4aa), AChR subunits (for example, CHRNAA1 and ACHE, and potassium channels (for example, KCNJ12) (2A). This pattern has been previously described only for the α subunit of the Na$^+/K^+$-ATPase, ATP1a2a, in E. electricus electric organs. Further, it was shown that Na$^+/K^+$-ATPase activity had different kinetic properties in each electric organ (16). Multiple α and β subunit isoforms of the sodium pump are expressed in the electric organs, the most abundant of which were ATP1a2a and ATP1b1a (see Fig. 2A and table S4).

Although the magnitude of change in abundance in comparing Hunter’s electric organ and Sachs’ electric organ was not as great as the comparison of main electric organ to Hunter’s/Sachs’ electric organs, most identified ATPase α and β subunits shared this pattern of main > Hunter’s > Sachs’, as did most AChR subunits, ACHE, and sodium channel proteins. Note that when the voltage-gated sodium channel was first isolated from E. electricus electric organ, only an α subunit was detected (no β subunit) (17). In the present analysis, we detected two β subunits (Fig. 2A), SCN4bα and SCN1Bb, both mid-range in terms of total abundance (table S2). Previous studies have localized two isoforms of the Na$^+/K^+$-ATPase α subunit in E. electricus electrocytes: one isoform present predominately on the innervated membrane face (an ATP1a1a type) and the other highly abundant on the noninnervated membrane face (ATP1a2a) (18), although the function of the ATPase at the innervated membrane face has yet to be elucidated. Here, we were able to quantify additional protein isoforms present in the three electric organs and muscle (a total of five α subunits and three β subunits) (Fig. 2A). They are less abundant than ATP1a2a (table S2), a finding that likely reflects the sensitivity of this
unbiased, quantitative “shotgun” proteomics approach. The general pattern of electrogenic proteins as most abundant in main electric organ and least abundant in Sachs’ electric organ was not observed with potassium channels (Fig. 2, A and B). Of the three potassium channel proteins we quantified, KCNA1 (Kv1.1) and KCNJ12 (Kir2.2) were most abundant in main electric organ, whereas KCNH6 (Kv11.1) was most abundant in Sachs’ electric organ. This was also an interesting finding because inwardly rectifying potassium current has previously been described as the dominant K⁺ current in electrocytes, whereas voltage-gated K⁺ current has been only weakly observed (19) or not observed at all (20–23). Our analysis reveals that we can detect voltage-gated K⁺ channels present in the electric organ at concentrations equal to or greater than those of the inwardly rectifying K⁺ channel (absolute abundance; Fig. 2B).

Global comparison of the proteome and phosphoproteome: Nonelectrogenic proteins that may play roles in electric organ development and/or function

Very little is known about the molecular differences among the three distinct electric organs in *E. electricus*. Microscopy revealed that electrocytes from main and Hunter’s electric organs were packed tightly next to one another, whereas electrocytes from Sachs’ electric organ were more separated from one another with abundant extracellular matrix (20). In addition, electrocytes in Sachs’ electric organ were larger in size than those in main electric organ and had larger invaginations (20). Further, action potentials from main and Sachs’ electric organs were similar in shape; but during each discharge, the action potentials occurred much more rapidly in main electric organ compared to Sachs’ electric organ—the mean action potential duration in main and Sachs’ electric

![Graphs showing proteome and phosphoproteome comparisons](image-url)

**Fig. 2.** Proteins important for electric organ discharge in *E. electricus* show distinct patterns of abundance across electric organs. (A) Relative abundance of electrogenic ion transporters in electric organs and muscle. The general pattern observed among most ion transporters (main > Hunter’s > Sachs’) is broken by a voltage-gated potassium channel. (B) Total abundance of potassium channels detected in this study. Values shown are log₂-normalized summed reporter ion intensity values in table S2.
organ electrolytes was 1.5 and 2.2 ms, respectively (17, 20). We hypothesized that, given the difference in energetic demands on the strong versus weak electric organs (intermittent, strong voltage versus continuous, weak voltage), there may be many underlying differences in protein abundance and protein phosphorylation that contribute to these distinct functional roles in *E. electricus*.

To tease apart these possible differences, we used a variety of clustering approaches. First, we clustered all quantified proteins and phosphopeptides in both fish (Fig. 3, A and B, respectively) to see how the electric organs and muscle relate broadly to one another and found that muscle is most distinct from all tissues. Of the three electric organs, Hunter’s and Sachs’ cluster most closely together, with main electric organ being most distinct. According to the degree of overlap among the top 10% most abundant quantified proteins and peptides in both fish (Fig. 3, C and D), protein and phosphopeptide levels showed that Hunter’s electric organ was the least distinct of the four tissues tested because it had the fewest unshared entries. At a protein level in both fish (Fig. 3C), the largest shared group in the Venn diagram was proteins shared among all three electric organs and muscle. Distinct tissue specificity was evident at a phosphopeptide level in both fish (Fig. 3D), with the largest shared group in the Venn diagram belonging to phosphopeptides common among the three electric organs only and not muscle. Thus, there are distinct protein phosphorylation patterns occurring in the electric organs compared to muscle.

We used k-means clustering to bin proteins and phosphopeptides that had at least twofold difference in abundance between the highest-expressing and lowest-expressing electric organ (Fig. 3, E and F; see summary of findings discussed in table S4). Clusters of unique proteins and phosphopeptides highly expressed in either main electric organ, Sachs’ electric organ, or muscle/Hunter’s electric organ emerged, illustrating that there are differences in protein abundance and phosphorylation among the three electric organs. We focused on these bins of differentially abundant proteins and phosphopeptides, notably the main electric organ and Sachs’ electric organ bins, which represent groups of proteins important for intermittent, high-voltage (main electric organ) or continuous, low-voltage (Sachs’ electric organ) discharges.

**Proteins and phosphopeptides highly abundant in main electric organ**

Several proteins important for electric organ discharge in the main electric organ protein cluster were identified (Fig. 3E, cluster 2, and table S4), including voltage-gated Na+ channel α and β subunits (SCN4aa, SCN4ab, and SCN1b), AChR subunits (CHRNA1), AChE, and the abundant β subunit for the Na+/K+–ATPase (ATP1b2). In addition, we observed calmodulin in the main electric organ–specific protein cluster, which has been previously described (11) to be very abundant in the electric organ from *E. electricus*, but differential abundance across electric organs has not been previously shown. We also found several electric organ discharge–related proteins in the main electric organ phosphopeptide clusters (Fig. 3F, cluster 3, and table S4), including AChE, AChR subunits (CHRNB1 and CHRNA1), ATP1a2a, the voltage-gated Na+ channels (SCN4aa and SCN4ab), and also a chloride channel (CLCN1a) that was not observed in the proteome analysis.

In addition to proteins involved in electric organ discharge, we also observed some other interesting distinctions with the main electric organ cluster. First, in the main electric organ–specific protein cluster (Fig. 3E, cluster 2, and table S4), we found an α-AMP–activated protein kinase (AMPK) subunit (PRKAA2, a catalytic subunit of AMPK). AMPK is the major energy sensor in cells (24). In addition, we found protein phosphatase 1aa, which can dephosphorylate the α subunit of AMPK (25) and falls in several other signaling pathways. We also identified several other proteins involved in other major signaling pathways, in particular mitogen-activated protein kinase (MAPK) and nuclear factor xB (NFκB) signaling pathways, which overlap significantly. For example, in the main electric organ phosphopeptide cluster (Fig. 3F, cluster 3), we identified phosphorylated amino acids within proteins in the MAPK cascade, including the MAPK proteins (MAP4K4, MAP3K5, and MAP2K2b), and the microtubule-associated protein tau (MAPTA and MAPTB). At a protein level, the transcription factor NFκB2, which is a final step in the MAPK signaling pathway, is highly abundant in main electric organ relative to Sachs’ electric organ, but it was not included in this initial clustering because it was not detected in one of the muscle biological replicates.

In addition, we identified several transcription factors in the main electric organ protein clusters, specifically CBX6b, a component of the polycomb repressive complex. It has been demonstrated that the polycomb repressive complex 1 is involved in maintaining transcriptional repression by chromatin/epigenetic mechanisms (26). We also identified ZNF644b (a predicted zinc finger transcription factor), mutations of which are associated with myopia in humans (27). In the main electric organ phosphopeptide cluster, we identified the transcription factor grainyhead-like 2a (GRHL2a) that is a paralog of upstream binding protein 1 (UBP1), which was abundant in Sachs’ electric organ (see below). Grainyhead family transcription factors have important and diverse roles in development, notably in the regulation of genes important for cell junction formation (for example, tight junctions and others) (28), which are vital to the development or establishment of polarization in cells (29), a property that is at the heart of the nature of electrolytes required for generating lethally high voltages. We also found an aquaporin protein (AQP8a.2) in the main electric organ proteome cluster, which is consistent with our underlying hypothesis that main electric organ is less energetically conservative compared to Sachs’ electric organ and must deal with the repercussions of massive ion flux (and, thus, osmotically driven water flux) for each action potential. Finally, one of the most abundant phosphopeptides belonged to NDRG4. In total, we quantified 17 phosphosites, 7 of which have not been identified previously in mammals (table S1) and 4 of which fell within the main electric organ phosphopeptide cluster. The function of this protein in electrolytes is not known.

**Proteins and phosphopeptides abundant in Sachs’ electric organ**

We found several interesting proteins in the Sachs’ electric organ–specific protein and phosphopeptide clusters [Fig. 3, E (cluster 1) and F (cluster 2), respectively]. We again found several proteins in numerous signaling pathways abundant in Sachs’ electric organ including ULK1, a serine/threonine protein kinase that (together with AMPK, the catalytic subunit of which is highly abundant in main electric organ; see previous section) is important for regulating cell autophagy (24). We also identified a Rho guanosine triphosphatase–activating protein (ARHGA21a), two guanine nucleotide exchange factors (ARHGEF1b and ARFGEFI), the transcription factors STAT3 and UBP1, a grainyhead family transcription factor, and a paralog of GRHL2a (highly abundant in main electric organ; see previous section). It has been demonstrated that UBP1 expression determines intercalated cell development and polarity, notably determining in which part of the membrane the H+–ATPase is localized (30).

The most abundant phosphopeptides in Sachs’ electric organ (Fig. 3F, cluster 2) belonged to proteins that are implicated in muscular diseases.
These include ANO5b, a protein in which mutations have been demonstrated in muscular dystrophy (31) and dilated cardiomyopathy (32). Although not detected in the whole-proteome experiment, at an RNA level, ANO5b shows increased abundance in Sachs’ electric organ over muscle, main electric organ, and Hunter’s electric organ at the RNA level (table S5), and its expression in this table indicates that it has a similar expression level to cardiac muscle. Also highly abundant in Sachs’ relative to other electric organs include CMYA5 (cardiomyopathy-associated 5), which, together with desmin, is thought to coordinate specific kinases and phosphatases in muscle (33).
**Phosphopeptides that show differential abundance compared to protein abundance in electric organs**

To determine whether specific phosphosites showed differential abundance that was independent of changes in protein abundance [defined as a difference between the log2 (tissue/median) phosphopeptide and protein values that was at least 1], we first examined the phosphosites found in proteins known to be important for electric organ discharge, that is, the AChR, ACHE, voltage-gated sodium channels, and the Na+/K+ ATPase. In general, most of the phosphopeptides in these proteins showed similar abundance to that of the protein quantitation (tables S3 and S6). In the most abundant ATPase (ATP1a2a), three fully tryptic phosphopeptides (representing three phosphosites) showed differential relative abundance, one of which appeared increased in abundance in main electric organ #2 (from biological replicate two, but not one) and two of which appeared reduced in abundance in main electric organ #2 compared to the protein-level abundance while remaining unchanged in Sachs’ electric organ (Fig. 4 and table S7). The location of these three phosphosites has been previously described in mammals; however, we were unable to find descriptions in the literature of the functions of these three phosphorylation sites.

In the voltage-gated Na+ channel protein SCN4aa, we identified four fully tryptic phosphopeptides (representing four phosphosites) that showed differential abundance compared to the protein abundance; three of these showed reduction in at least one Sachs’ electric organ sample (with no change in other electric organs) (Fig. 4 and table S7). These distinct phosphosites localized with 100% probability (see search methods) and are not previously described in humans (9); one of these (AVSHASFLSQLK) falls within the C-terminal domain of SCN4aa. The fourth phosphoprotein showed reduction in main electric organ #2 (with no change in other electric organs) and, despite being not fully localized, represents a novel phosphorylated region because it is not near a known phosphosite in mammals (9).

Finally, we identified two phosphosites in ACHE, both of which have not been described in mammals and were fully localized to a specific amino acid (see Materials and Methods and table S7). One of these phosphosites showed a large decrease in Hunter’s electric organ from fish #2 that was not reflected in its protein concentration, and showed slight decreases in main electric organ #2 and slight increases in Sachs’ electric organ #2. The second phosphosite showed a decrease in both Sachs’ electric organs (up to approximately twofold). Future studies will be needed to provide additional biological replicates, confirm the stoichiometry of these phosphorylation sites, and shed light on the biological importance of these particular phosphosites in different electric organs.

**Tissue-specific abundance of desmin phosphorylation**

Differential abundance of desmin protein isoforms and protein phosphorylation was previously reported in the three electric organs of *E. electricus* (34, 35). We identified two desmin proteins, desmin A and desmin B, which are similarly abundant in all electric organs at a whole protein level. However, when we considered phosphopeptides, we observed that distinct phosphopeptides were more abundant in one tissue than those in another tissue. A total of 70 phosphopeptides for desmin A and 8 for desmin B were observed. Of these, there are two phosphopeptides for desmin A that fall under phosphopeptide cluster 1 (high main electric organ) and two under cluster 4 (high Sachs’ electric organ), and for desmin B, there is one phosphopeptide that fall under cluster 1 (high main electric organ) and one under cluster 3 (muscle cluster; highest in Hunter’s electric organ, variable in Sachs’ electric organ, and very low in main electric organ). These data agree with the previous reports showing differential protein and phosphorylation abundance across the three electric organs (34, 35) and provide locations of phosphorylated residues in these proteins.

The function of desmin in electrocytes is unknown, as is the function of desmin, generally, especially in fish (33). Desmin is an intermediate filament protein, but roles for desmin extend beyond structural: It is involved in calcium homeostasis (33), is abundant in neuromuscular junctions and at the points of attachment between cardiomyocytes (36), is important for mitochondria structure and function, and can regulate the expression of genes important in myogenic and cardiogenic regulation (37). Mutations in desmin are associated with human disease, and various posttranslational modifications including protein phosphorylation have been implicated in desmin-related disease (37).

**Proteins involved in neuromuscular development largely absent in proteomic and phosphoproteomic analysis**

Given the important historical role electric fishes (especially *Torpedo* species) have played in understanding neuromuscular junction formation and function, we considered whether there were unique patterns of protein abundance or phosphorylation in proteins important in neuromuscular junction development. We focused specifically on proteins involved in the early stages of development: prepatternning of AChRs in the future postsynaptic membrane before the arrival of a nerve axon and early signaling events in synapse formation after axon arrival.

The process of neuromuscular junction formation is well conserved in animals. Before the arrival of the innervating neuron, AChRs are pre-patterned into clusters on the future postsynaptic membrane [reviewed by Wu et al. (38)]. Prepatternning in zebrafish occurs slightly differently compared to mammals. In mammals, Lrp4 is required for prepatternning but not in zebrafish, whereas Wnt4a and Wnt11r are required for prepatternning in zebrafish but not in mice (39). Also required for ACHR prepatternning in zebrafish and mammals is the protein rapsyn, which interacts with and regulates AChR prepatternning [reviewed by Wu et al. (38)]. During the process of synapse development, agrin is expressed from the extending motoneurons (38). As the axons approach developing muscle, agrin binds to Lrp4 on the muscle; the binding of neuronal agrin to muscular Lrp4 causes Lrp4 to associate with the kinase MuSK (40). This association causes MuSK to become phosphorylated [first observed in *T. californica* (41)], triggering a cascade of responses that promote the development of the neuromuscular junction (40).

We identified genes for *agrin, rapsyn, musk, wnt4a*, and *wnt11r* in our genome, whereas *lrp4* was not annotated in our genome. Attempts at identifying a misannotated *lrp4* in our genome failed; we hypothesize that this may be because *lrp4* falls within a region of the genome that is difficult to assemble. An investigation at the zebrafish genome database at Ensebm and ZFIN revealed that *lrp4* is not localized within the zebrafish genome, perhaps reflecting a difficult genomic region to assemble that is conserved in both *E. electricus* and zebrafish. Of these five genes that we identified in our genome assembly, all were quantified at an RNA level, but only agrin and rapsyn were detected in our proteome analysis and none was detected in our phosphoproteome analysis.

The reason for our lack of identifying these key proteins in our analysis is unclear. We checked the length of these five protein sequences against that of zebrafish to determine whether we failed to identify a subset of them because of gene model issues, but they appear to be at or very near to full length. Further, we checked whether we identified peptide spectra that matched these proteins but nonuniquely such that they were not used in quantitation (see raw output from Proteome Discoverer and tables S8 and S9), but that also is not the case. Because
the MS analyses identify the most abundant ions in time at the cost of not identifying more lowly abundant species, this could be a consequence of other proteins being extremely abundant in the electric organ of *E. electricus*.

**Muscle-specific proteins more abundant in Hunter’s electric organ compared to other electric organs**

It is interesting to note that in the *k*-means clusters (both at the whole-proteome and phosphoproteome resolution), Hunter’s electric organ consistently expresses proteins intermediately between main and Sachs’ electric organs and also more similarly to muscle compared to main and Sachs’ electric organs. This was not revealed in the hierarchal clustering analysis (Fig. 3, A and B) in which both Sachs’ and Hunter’s electric organs clustered more closely to muscle than main electric organ; this is likely because in the hierarchal clustering analysis, all quantified proteins were included, whereas within the *k*-means clustering analysis, only a subset of proteins or phosphopeptides with at least twofold abundance difference among the electric organs were included. This observation is of interest since we used the strong-voltage portion of Hunter’s electric organ (the section under main electric organ) for these analyses, and morphologically, Hunter’s electric organ...
appears most similar to main electric organ (20, 42), although no studies have comprehensively characterized or compared electric organs.

**Correlation of mRNA and protein abundance**

Recent research efforts have focused on mRNA sequencing to understand the repertoire of genes expressed in the electric organs of *E. electricus* (6, 43), but it is not certain that these measurements always reflect the actual changes in protein abundance. To determine whether the protein abundance values correlated with mRNA expression values, we compared log2 (electric organ/muscle) ratios for both the mRNA expression and protein abundance levels of the expressed products from the 2866 gene models for which we obtained protein abundance data. Our results (Fig. 5, A and B, and table S10) show that most of the gene models are not differentially expressed at either an mRNA or protein level. Similar numbers of gene models are differentially expressed only at the protein level, only at the RNA level, or at both the RNA and protein level, in any given electric organ. The fewest and perhaps most intriguing gene models are those that are differentially expressed at both the RNA and protein levels but in opposite directions.

For example, in main electric organ, we saw a putative FXYD protein (gene model scaffold12.g65) displaying discordant RNA and protein abundance values relative to muscle. FXYD proteins are very small proteins that copurify with the Na+/K+-ATPase, but for which a clear function has not yet been elucidated (44). At the mRNA level in main electric organ, the FXYD gene expression appears down-regulated compared to muscle. However, at the protein level, it appears up-regulated. A similar result was observed for Hunter’s electric organ, but the results were less marked than those of main electric organ. In Sachs’ electric organ, differential abundance at the protein level (but not mRNA level) was observed. In the quantitative proteomic experiment, this FXYD protein is highly abundant in main electric organ compared to Sachs’ or Hunter’s electric organs only in one of the biological replicates (table S2) and is of similar abundance across all three electric organs in the other biological replicate. It is important to note that at an mRNA level, this FXYD transcript is not lowly expressed; its abundance is very high in muscle and quite high in the three electric organs. In main electric organ, we also observed that the ATP1a3a (an α subunit of Na+/K+-ATPase) protein and RNA abundance ratios are discordant (highly abundant in main/muscle at a protein level and lowly abundant at an RNA level), although it does not appear differentially expressed at either the protein or RNA level in Sachs’ or Hunter’s electric organs.

**DISCUSSION**

Electrocytes found in the electric organs of *E. electricus* are large, axially flattened cells with two major membrane faces: the posterior, innervated membrane face, which receives nervous stimulation and contains numerous sodium channels, and, opposite to it, the noninnervated membrane face, which is highly invaginated and contains numerous Na+/K+-ATPase. This polarized arrangement of electrocytes within electric organs is critical for the small individual voltages generated in each electrocyte to summate and create the larger voltages that can be emitted [reviewed by Gotter et al. (17); Fig. 1A]. Upon activation by acetylcholine release from the innervating neuron, AChRs open, allowing Na+ to flow into the cell and causing a slight membrane depolarization. This slight depolarization of the membrane, in turn, causes the voltage-gated Na+ channels to open, and a massive Na+ influx across the innervated membrane face occurs, driving the innervated membrane face into positive potentials. Because the noninnervated membrane face remains at negative potentials due to the abundance of Na+/K+-ATPase and the absence of Na+ channels in the membrane, there is a large transcellular potential difference that is achieved within each electrocyte, and because electrocytes are arranged massively both in series and in parallel, voltages and the current generated in each electrocyte summate. Na+/K+-ATPase and potassium channels reset the system and return the membranes to the “resting” membrane potential.

Electric organs are uniquely dedicated to generating and emitting endogenous voltages; therefore, among all abundant proteins in these organs were those vital to initiating the electric organ discharge (nicotinic AChRs, carrying mostly Na+ current), propagating the action potential (voltage-gated Na+ channels), and returning cells to and maintaining homeostasis (ACHE, Na+/K+-ATPase, and K+ channels). Our untargeted MS-based analysis measured deeply into the proteome of electrocytes, but because these six proteins are vital for electric organ discharges in *E. electricus*, we paid particular attention to their relative protein and phosphopeptide abundances among the three electric organs. Given the functional differences the distinct electric organs play in the *E. electricus*, specifically generating strong voltages intermittently in predation and defense or weak electric organ discharges in navigation and communication, we hypothesized that we would observe differences in protein abundance or phosphorylation reflecting these different functions.

Our analysis revealed several novel and known phosphorylation sites on electrogenic proteins, including the Na+/K+-ATPase (ATP1a2a), the voltage-gated sodium channel (SCN4aa and SCN3ab), the nicotinic AChR subunits (CHRNB1, CHRNDD, and CHRNFG), and ACHE. Notably, there was one phosphosite detected on both CHRNDD and CHRNFG that has been previously reported only in the electric ray species *T. californica*; the detection of this phosphorylation site only in electric fish species could represent a specific adaptation of electric fish proteins or could be a consequence of the AChR being much more abundant in electric organs compared to nonelectric tissue, facilitating their detection in electrogenic tissue; we did not identify any phosphosites on the voltage-gated potassium channels. These five protein families play fundamental roles in the biology of muscle and nerve cells. Thus, the identification of phosphorylation sites in the electric organs that have not been previously seen in studies with nonelectric animals may suggest that either these sites play a unique role in the generation of electric discharges seen only in electric fish or their unique abundance in the electric organ has allowed us to identify sites that are not apparent in studies with other animal tissues.

Our analysis identified several novel phosphorylation sites in our electrogenic proteins, including four novel sites within the C-terminal regulatory domain of SCN4aa (fig. S1). Using protein alignments of the C-terminal domain in several species, we found that the first novel phosphosite fell at phosphorylatable residues in all electric fish tested (both Gymnotiformes and one mormyrid) and a subset of other nonelectric fish and humans. Perhaps, our ability to detect this protein in *E. electricus* despite it not being previously detected in humans is due to its shear abundance in electrocytes. The second phosphosite was at phosphorylatable residues in *E. electricus* and a phosphomimetic D or E in all other species tested (electric or otherwise); this finding may be indicative of the importance of this residue in all species but an increased requirement of regulation by phosphorylation in the context of a strong-voltage organism. A third phosphosite was at phosphorylatable residues found only in *E. electricus*, whereas a fourth phosphosite was either at phosphorylatable residues or a phosphomimetic D in Gymnotiformes only, indicating the possibility of unique protein regulation within the only strong-voltage gymnotiform or all the Gymnotiformes,
respectively. Together, these findings indicate that there may be unique functional adaptations within the C-terminal domain of SCN4aa in *E. electricus* and other electric fishes. Further experimentation is required to test each of these hypotheses. Notably, protein crystal structures with the C-terminal domain of SCN4a, calmodulin (historically noted as being highly abundant in *E. electricus* electric organs), and FGF13 have highlighted that SCN4a can be regulated by calmodulin and FGF13 via this domain; this is interesting because we identified 10 phosphosites in *E. electricus* FGF13, and an interesting pattern of abundance of calmodulin across our electric organs being one of the most abundant proteins in strong-voltage electric organ (main and the strong-voltage section of Hunter’s electric organs) and much less abundant in the weakly electric Sachs’ electric organ and muscle. Together, this may indicate a potential regulatory mechanism of the voltage-gated Na⁺ channel via phosphorylation in the C-terminal domain and interactions with calmodulin and FGF13; future experiments are required to test this hypothesis.

We also identified a unique and consistent abundance pattern in several electrogenic proteins, including several Na⁺/K⁺-ATPase, several voltage-gated Na⁺ channel subunits, two K⁺ channels (voltage-gated and inwardly rectifying), AChR subunits, and ACHE (Fig. 2A), that show the highest abundance in main electric organ, lowest abundance in Sachs’ electric organ, and intermediate abundance in Hunter’s electric organ. Since the influx of sodium is energetically costly and there are very large fluxes of sodium required for the amount of power consumed in the lethal strong-voltage discharge emitted from main and Hunter’s electric organs compared to the weak electric discharge of Sachs’ electric organ, our observation that Sachs’ electric organ has a lower concentration of the two sodium channels and the sodium pump is not unexpected. Thus, the differences in protein abundance of ion channels and pumps across the three electric organs seem to reflect different demands on the cells found in each electric organ and a trade-off between continuously generating low-voltage discharges while minimizing energetic costs and intermittently generating high voltage at a higher energetic cost.

**Fig. 5. Correlation of RNA expression and protein abundance.** (A) Plots showing the correlation of RNA expression and protein abundances for all proteins detected in unenriched proteomics experiment. See the “Correlation of RNA expression and protein abundance values” section for details on how graphs were generated. DE, differentially expressed genes or proteins. (B) Table showing counts of gene models falling in each category depicted on graphs in (A).

---

### Table: Correlation of mRNA expression and protein abundance

<table>
<thead>
<tr>
<th>Observation</th>
<th>Eel specimen #1</th>
<th>Eel specimen #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main 1</td>
<td>Sachs’ 1</td>
</tr>
<tr>
<td>(1) No DE in RNA or protein</td>
<td>1840</td>
<td>1904</td>
</tr>
<tr>
<td>(2) DE in protein only</td>
<td>399</td>
<td>361</td>
</tr>
<tr>
<td>(3) DE in RNA only</td>
<td>220</td>
<td>253</td>
</tr>
<tr>
<td>(4) DE in both RNA and protein</td>
<td>255</td>
<td>223</td>
</tr>
<tr>
<td>(5) DE in RNA and protein, opposite direction</td>
<td>36</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Main 2</th>
<th>Sachs’ 2</th>
<th>Hunter’s 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) No DE in RNA or protein</td>
<td>1778</td>
<td>1886</td>
<td>2134</td>
</tr>
<tr>
<td>(2) DE in protein only</td>
<td>440</td>
<td>360</td>
<td>145</td>
</tr>
<tr>
<td>(3) DE in RNA only</td>
<td>242</td>
<td>316</td>
<td>346</td>
</tr>
<tr>
<td>(4) DE in both RNA and protein</td>
<td>232</td>
<td>145</td>
<td>89</td>
</tr>
<tr>
<td>(5) DE in RNA and protein, opposite direction</td>
<td>31</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

---


10 of 16
Remarkably, the pattern of electrogenic proteins being highest in main electric organ, lowest in Sachs’ electric organ, and in the middle for Hunter’s electric organ was broken by an additional voltage-gated K⁺ channel (KCNH6; Fig. 2, A and B). This was noteworthy because, first, it is most abundant in Sachs’ electric organ and, second, to our knowledge, voltage-gated K⁺ channel proteins have not been confirmed in *E. electricus* electric organ previously. One patch-clamp study of the innervated membrane of *E. electricus* electocytes revealed small, low-density, voltage-gated outward K⁺ currents in a subset of patches examined (19), despite previous studies failing to detect such a current (20–23). Expression of an mRNA transcript for KCNA1 in the electric organ from *E. electricus* was confirmed later and encodes a functional protein when expressed in a heterologous system (45). The channels might be under continuous inhibition that is relieved under certain physiological conditions (45) or be trafficked to the membrane in a circadian fashion or due to social cues, as observed with ion channels in the weakly electric *S. macurus* (46). Although we have evidence that these potassium channel proteins are translated and of reasonable abundance (and in greater total abundance than KCNJ12 in some of the electric organs; Fig. 2B), the function of these proteins in the electric organs remains elusive.

Finally, to more comprehensively compare the difference in protein abundance and phosphorylation among the three electric organs, we used two clustering approaches. We observed that main electric organ consistently clusters separately from Sachs’ and Hunter’s electric organs (Fig. 3, A, B, E, and F). We also observed that although main and Sachs’ electric organs have their own distinct clusters of highly abundant proteins and phosphopeptides, this is not true for Hunter’s electric organ, which has no highly abundant clusters of its own, clusters more closely to Sachs’ electric organ than to the other strong-voltage organ (main), and seems to share its most highly abundant proteins with either muscle or main electric organ. This was a surprise to us because we had assumed going in to this experiment that the strong-voltage section of Hunter’s electric organ would likely look very similar to main electric organ. Perhaps, this indicates a developmental or evolutionary history of the three distinct electric organs of *E. electricus* specifically that remains to be elucidated.

Our clustering approach also revealed several interesting groups of proteins or phosphopeptides highly abundant either in main electric organ or in Sachs’ electric organ that included transcription factors, proteins involved in signaling pathways, and proteins involved in cellular energetics. Phosphopeptides for the protein NDRG4 were consistently among the most abundant phosphopeptides detected in main electric organ, and four of them fell in the main electric organ–specific phosphopeptide cluster. The function of this protein is not well understood, but in humans, NDRG4 mutations are associated with elongated QT interval and other heart defects (47–49). In zebrafish, knocking down the expression of NDRG4 gives rise to fish with heart defects, including reduced heart rate, which is interesting, as treatment of zebrafish with QT-prolonging drugs results in decreased heart rate (47). Because NDRG4 is important to regulate the rate of repolarization of cardiomyocytes, perhaps, it is important to regulate repolarization of electocytes because the action potentials in main electric organ are shorter than in Sachs’ electric organ (20). We also identified two paralogous transcription factors, GRHL2a and UBP1, that are abundant in main and Sachs’ electric organs, respectively. Both GRHL2 and UBP1 have been demonstrated in model system to be important in the development of cell junctions and the establishment of cell polarity (28–30). Although these studies are in model systems and not in *E. electricus*, perhaps, UBP1 in Sachs’ electric organ is performing a similar but distinct function as GRHL2 in main electric organ in the development or maintenance of cell polarity, but the differences in protein isoform expressed in each tissue reflect specific and distinct requirements of each cell type.

Together, these data represent the first quantitative proteomic and phosphoproteomic description and comparison of electric organs in any electric fish species. The findings in this study suggest key differences among the three electric organs in *E. electricus*. Sachs’ electric organ appears to be more energetically conservative than strong-voltage main electric organ, by expressing fewer Na⁺ channels, and thus requiring less adenosine 5’-triphosphate hydrolysis by having fewer Na⁺/K⁺-ATPases to maintain resting membrane potential. This finding is consistent with the key functional differences between Sachs’ and main electric organs; Sachs’ electric organ is used continuously for navigation and communication, whereas main electric organ is used only intermittently for strong-voltage discharges in predation and defense. We also found differences among Sachs’ and main electric organs that reflect the differences in signal transduction, as well as transcription factors that may be important in tissue form and function, given their known roles in establishing cell polarity and cell-cell contacts. We identified interesting patterns of potassium channel expression related to two voltage-gated potassium channels, one of which is expressed most abundantly in Sachs’ electric organ. This finding is of interest because voltage-gated potassium currents have been shown to be very minor currents in Sachs’ electric organ, when detected at all. Our observations concerning the voltage-gated K⁺ detected in this study suggest the need to determine the function of these channels in electocytes and why they are so abundant. Finally, this study describes several new phosphorylation sites in proteins important for electric organ discharge. Given the large role that sodium channels and pumps play in brain function and overall human physiology, a greater understanding of how these proteins are regulated at a translational and posttranslational level in electric fishes may provide important insights into previously unrecognized means of regulation in behavioral and physiological functions in humans and other animals.

**MATERIALS AND METHODS**

**Genome assembly improvement and gene prediction**

**Genome scaffolding and gap filling**

To improve on the existing *E. electricus* genome assembly and the downstream gene model prediction, additional raw sequencing reads were incorporated into the publicly available genome assembly (termed “SOAPdenovo2assembly”). Scaffolding of the SOAPdenovo2 assembly was performed using SSPACE (Standard v3.0) (50), incorporating both Illumina 2.5-kb insert mate pair (MP) libraries and 454 libraries, with the command -b -z 500 -g 3 -v 1 -T 27 and following library information given in the library file (.l1): Bowtie was designated for all libraries (column 2), and in columns 5 to 7, “2500, 0.2, RF” was designated for Illumina MP reads and “12000, 333, RR” for 454 reads.

Following scaffolding, gap closing was performed with GapCloser (v1.12) (51) using both short (2 × 100) Illumina paired-end reads and Illumina MP reads (2.5-kb insert), specifying “max_rd_len=120” globally in the configuration file, and the following reads in each library type. For Illumina MP reads, “avg_ins=3000, reverse_seq=1, asm_flags=4, rd_len_cutoff=60, rank=1, pair_num_cutoff=5, map_len=35.” For Illumina paired-end reads, “avg_ins=230, reverse_seq=0, asm_flags=4, rd_len_cutoff=80, rank=1, pair_num_cutoff=3, map_len=32.” This scaffolded and gap-closed assembly will be referred to later in this text as “GapClosed Assembly.”
Gene prediction and annotation

For the purpose of gene prediction, RNA sequencing reads from eight tissues of *E. electricus* [brain, spinal cord, heart, skeletal muscle, main electric organ, Sachs' electric organ, Hunter's electric organ, and kidney; see Gallant et al. (6) and Traeger et al. (43)] were mapped to the scaffolded and gap-filled genome with STAR (v2.4.0) ([Gallant et al.](https://advances.sciencemag.org/)) using default parameters. Alignments from all eight tissues were merged, and extrinsic hints for "intron" and "exonpart" features were generated using custom scripts from the STAR output "SJ.out.tab" files and "Aligned.out.sam" files, respectively.

In addition to transcript hints, we also incorporated protein evidence based on human proteins (Ensembl release GrCh37.73), zebrafish proteins (Ensembl release Zv9.73), and all other Gymnotiform protein sequences available at National Center for Biotechnology Information (NCBI) Taxonomy Browser (downloaded 7 February 2014). To do this, we followed the tutorial offered by the AUGUSTUS makers (http://bioinf.uni-greifswald.de/bioinf/wiki/pmwiki.php?n=Augustus.IncorporateProteins).

Briefly, all protein sequences were blasted against the genome assembly (tblastn –evalue=1e-15 –threshold=999 –max_intron_length=3000 –max_target_seqs=10). Top hits were parsed, and then Exonerate (v2.2.0) (53) was run on the top hits using default parameters. The output from Exonerate was converted to hints for AUGUSTUS. Protein hints were generated with scripts from K. Hoff (University of Greifswald and part of AUGUSTUS team, personal communication).

AUGUSTUS (v2.6) (54) was run with the following parameters: "--species=human --codingseq=on --gff3=on --alternatives-from-evidence=true --allow_hinted_splicesites=atac --UTR=on --uniqueGeneId=" with the following parameter setting in the extrinsic .cfg file:

```
start 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1e3
step ... 1 1 M 1 e+100 RM 1 1 L 1 1 W 1 1 P 1 1e2

orientation 1 1 1 0.997 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 1 2 1 0.96 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 1 3 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 1 4 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 1 5 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 2 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 3 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 4 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

orientation 5 1 1 M 1 e+100 RM 1 1 E 1 1 W 1 1 P 1 1

```

Annotation

Gene names were assigned to the AUGUSTUS gene models by comparison to *Danio rerio* Ensembl Zv9 build 79 protein sequences, in a method previously described (6). Briefly, this method assigns annotations stepwise based first on reciprocal blast to *D. rerio* proteins (same proteins match in *D. rerio* and *E. electricus* genome by reciprocal blast), second based on synteny of the *E. electricus* genome to the *D. rerio* genome, or finally top hit to *D. rerio* or nonredundant NCBI protein database, if the other two methods yielded no annotation. The method is described in full (6).

Analysis of improved genome and gene models

MS approaches rely on a reference proteome to match experimentally derived spectra to a corresponding protein sequence. To aid in maximally identifying mass spectra, we aimed to improve the existing *E. electricus* genome assembly (termed "SOAPdenovo assembly," for the purposes of this study) (6) by incorporating additional sequencing reads to further scaffold followed by a round of gap filling (termed "gap filled assembly"); see Materials and Methods). To compare the SOAPdenovo2 genome assembly to the assembly obtained after a round of scaffolding and gap filling, we used Quast (v2.3) (55) to generate statistics about the SOAPdenovo2 assembly, the assembly after scaffolding only, and the assembly after scaffolding and gap filling (table S11A). Compared to the previous SOAPdenovo assembly, the gap-filled assembly showed significant improvement, having significantly increased contig and scaffold lengths (the longest contig length is approximately seven times longer and the longest scaffold length is ~3.5 times longer in gap-filled assembly compared to SOAPdenovo assembly) and a reduced number of contigs and scaffolds (~3.5 times and 2.5 times fewer, respectively); that is, the new gap-filled assembly had more of the assembly in longer assembled pieces.

In addition, we wanted to determine whether the AUGUSTUS gene models predicted in this study were an improvement over the previously published gene models. To that end, predicted proteomes from each source were blasted against *D. rerio* proteins (Ensembl Zv9 build 79) and compared to one another based on blast hit scores (table S11B). We used blastp with default parameters and extracted only top hits for every *E. electricus* protein along with their corresponding score. Then, using custom scripts, we compared the top score for each *D. rerio* protein ID. In this approach, we found that 6714 times the "old" and "new" gene models had identical blastp scores, 7365 times the "new" gene models had a better score, and 2619 times the "old" gene models had a better blast score. That is, we found that, ~84% of the time, gene models predicted in the gap-filled assembly were as good as or better than the previous gene models in the SOAPdenovo assembly. The *E. electricus* proteome database for mass spectral searching was derived from these new gene model predictions.

Calculation of RNA expression levels

The GFF output from AUGUSTUS was converted to GTF format using custom scripts for input into HTSeq (56). Expression values were calculated at the gene level (as opposed to transcript level) for every AUGUSTUS gene model using the previously described STAR-mapped RNA sequencing reads to genome and the htsq-covt command from HTSeq (“--intersection-strict -a 3 -t exon –s no -i gene_id”) (v0.6.1p1) (56), normalized for library size with DESeq (v1.18.0) (57), and further normalized for transcript length (exonic length) to generate "reads per kb transcript." These values are found in table S5.

Protein isolation and mass spectral analyses

Protein isolation and digestion

All animal care procedures used in this study followed the guidelines outlined by the American Physiological Society and were approved
by the Institutional Animal Care and Use Committee at the University of Wisconsin-Madison. Overview of the method used is depicted in Fig. 6. Two ~61-cm specimens were used in this study. Tissue was dissected, flash-frozen in 50-ml conical tubes, and stored at −80°C for later use. Protein isolation through protein digestion occurred on the same day to avoid freeze/thaw on intact proteins and was performed on main electric organ, Sachs’ electric organ, Hunter’s electric organ, and skeletal muscle from two fish using a methanol/chloroform extraction method. A small piece of each tissue was removed from the stock tissue and was pulverized in liquid nitrogen using a ceramic mortar and pestle. Five milliliters of urea lysis buffer was added [8 M urea, 25 mM tris (pH 8), 100 mM sodium chloride, 25 mM sodium fluoride, 10 mM sodium pyrophosphate, 50 mM β-glycerophosphate, and 1-Roche cOmplete EDTA-free Protease Inhibitor Cocktail Tablet and 1-Roche PhosSTOP Inhibitor Tablet per 20 ml of buffer]. Samples were homogenized using a Tissue Tearor for 1 to 2 min at 4°C. In a 50-ml conical tube, methanol/chloroform extraction was performed by adding 4 volumes of MeOH and 1 volume of CHCl₃, followed by vortexing; 3 volumes of H₂O were added, followed by vortexing and then spinning for 5 min at 4000 relative centrifugal force. After spinning, the top, aqueous phase was removed, leaving the interphase intact. Excess methanol was added, followed by vortexing and spinning for 5 min. The supernatant was removed, and the pellet was washed for a second time with excess methanol, vortexed, and centrifuged. Pellets were dried slightly and were resuspended in 8 M urea and 50 mM ammonium bicarbonate containing Roche PhosSTOP (1 tablet/20 ml). Protein concentrations were determined using bicinchoninic acid assay (BCA) (Thermo Fisher Scientific), and because BCA assays require lower urea concentrations, a small aliquot of the sample in 8 M urea was used and diluted down to 1 M urea. After protein quantitation, a total of 1.5 mg was diluted to 1.5 M urea in 50 mM ammonium bicarbonate containing Roche PhosSTOP, and incubated with dithiothreitol (DTT) at a final concentration of 5 mM for 35 min at 65°C. Alkylation was performed by incubating with iodoacetamide at a final concentration of 12.5 mM for 2.5 hours on a shaking table. Reactions were quenched by adding 24 µl of 5% hydroxylamine and incubating for 15 min at room temperature. To test mixing ratios, 5 µl of each sample was pulled from each sample and combined in a 1:1:1:1:1:1:1:1 ratio with the other samples. The remaining labeled peptides were frozen at −80°C until later use.

For the TMT test mix, samples were desalted using C18 columns (OMIX C18 100-µl tips, part number A57003100, Agilent Technologies) and run on an Orbitrap for ratio testing. Summed reporter ion intensities were used to normalize, finding first the sum of all reporter ion intensities for all peptides, finding the median summed value for all eight channels, and then using this to determine how to adjust the volume input to ensure equal ratios were added for all eight channels. **Prefractionation using high-pH reversed-phase chromatography**

Labeled samples (1 mg) were thawed, combined in equal ratios based on above calculations, and dried down to a volume of ~200 µl in a SpeedVac. Samples were brought up in ~1 volume of buffer A for high-pH reversed-phase chromatography. Separation was performed on a high-performance liquid chromatography (HPLC) (Waters 2795) fitted a Gemini 5u C18 250 × 100 column (Phenomenex) with a flow rate of 5 ml per minute. Buffer compositions were as follows: buffer A, 10 mM ammonium formate in H₂O (pH 10); buffer B, 10 mM ammonium formate in 80% ACN (pH 10). The following gradient was used for separation: 0 to 2 min, 100% buffer A; 2 to 3 min, 0 to 5% buffer B; 3 to 23 min, 5 to 60% buffer B; 23 to 25 min, 60 to 100% buffer B; 25 to 26 min, 100% buffer B; 26 to 27 min, 0 to 100% buffer A; 27 to 35 min,

---

**Fig. 6. Overview of experimental method.** Flowchart showing overview of procedures in this study.
100% buffer A. In total, 70- to 30-s (2.5 ml) fractions were collected. These fractions were combined to 1-min fractions. For whole-proteome analysis, 500 μl of each 1-min fraction was removed, and 4.5 ml was retained for phosphopeptide enrichment. Samples (500 μl) were dried in a SpeedVac; 4.5 ml of samples was lyophilized. All samples were stored at −80°C after drying.

**Unenriched, whole-proteome data acquisition**

For whole-proteome analysis, fractions (see above section) were combined in the following manner: From 1-min fraction #5 (combined 30-s fractions #9 and #10) to 1-min fraction #31, every sixth fraction was combined such that six final samples were obtained. Peptides were analyzed by nanoLC-MS/MS using the Agilent 1100 nanoflow system (Agilent Technologies) connected to a new-generation hybrid linear ion trap-orbitrap mass spectrometer (LTQ-Orbitrap Elite, Thermo Fisher Scientific) equipped with an EASY-Spray electrospray source. Chromatography of peptides before mass spectral analysis was accomplished using capillary emitter column (PepMap C18, 3 μm, 100 Å, 150 × 0.075 mm, Thermo Fisher Scientific) onto which extracted peptides were automatically loaded. NanoHPLC system delivered solvents A [0.1% (v/v) FA] and B [99.9% (v/v) ACN and 0.1% (v/v) FA] at 0.50 μL/min to load the peptides (over a 30-min period) and 0.3 μL/min to elute peptides directly into the nanoelectrospray with gradual gradient from 3% (v/v) B to 30% (v/v) B over 154 min, followed by 10-min fast gradient from 30% (v/v) B to 50% (v/v) B, at which time a 7-min flash-off from 50 to 95% (v/v) B took place. As peptides eluted from the HPLC column/electrospray source, MS scans were acquired in the Orbitrap with a resolution of 120,000 followed by HCD-type MS2 fragmentation at 30,000 resolving power of 10 most intense peptides detected in the MS1 scan from 400 to 2000 mass/charge ratio (m/z); redundancy was limited by dynamic exclusion. Activation settings for the optimum TMT-based MS2 fragmentation were as follows: default charge, 2; isolation width, 2 m/z; normalized collision energy, 38; and activation time, 0.1 ms. Raw data were directly imported into Proteome Discoverer (v1.4.1.14), where protein identifications and quantitative reporting were generated (see the “Database development, searching, and FDR estimation” section for search details).

**Phosphopeptide enrichment by titanium dioxide chromatography**

For phosphopeptide enrichment, fractions were combined in the following manner: From 1-min fraction #5 (combined 30-s fractions #9 and #10) to 1-min fraction #31, every fourth fraction was combined such that four final samples were obtained. We found that despite being in a volatile salt solution, at these volumes (a total fraction volume of ~31 ml per combined sample), these resulting samples remained too salty for the titanium dioxide chromatography method, so each fraction was desalted on a 3-cm² C18 column. The desalted samples were then enriched for phosphopeptides using titanium dioxide chromatography, in a method based on the work of Sugiyama et al. (58), but with the following modifications: Titanium dioxide tips were prepared with 1.5 mg of titanium dioxide (10 μm) per column, packed atop a C8 disk. Desalted fractions were dried to completion and resolubilized in 100 μl of loading buffer (1 M glycolic acid, 80% ACN, and 5% TFA). The column was washed twice with loading buffer, and then the sample was loaded, washed twice with 100 μl of loading buffer, and washed twice more with 100 μl of wash buffer (80% ACN and 1% TFA). Enriched phosphopeptides were eluted twice from the titanium dioxide with 50 μl of 1% ammonium hydroxide and eluted from the C8 disk with 30% ACN and 0.1% TFA. The eluate was acidified with 3.5 μl of neat FA and dried using a SpeedVac.

**Phosphoproteome data acquisition**

Phosphopeptides were analyzed on the new Orbitrap Tribrid Model Fusion Mass Spectrometer fitted with a nanoflow LC column, as previously described (59).

**Database development, searching, and FDR estimation**

For mass spectral searching, an E. electricus protein database was derived from AUGUSTUS gene models (see previous section). Because often a single-gene model had more than one possible coding sequence predicted as a reflection of alternative splicing predictions, to avoid losing quantitative power and still retain the limitation of only quantifying uniquely mapping peptides, all protein sequences predicted for a single-gene model were concatenated such that they represented a large, single sequence in the .fasta file. This input .fasta file was used as the protein database for Proteome Discoverer.

For our untargeted, whole-proteome analysis, database searching was performed with the following parameters: Thermo Proteome Discoverer (v1.4.1.14), using the Sequest HT search engine platform to interrogate the E. electricus database (see above) plus common laboratory contaminants. Cysteine carbamidomethylation and TMT-specific labeling were selected as static modifications, whereas methionine oxidation and asparagine/glutamine deamination were selected as dynamic modifications. Peptide mass tolerances were set at 0.05 ppm for MS1 and 0.02 Da for MS2. A decoy database of reverse sequences was used for false discovery rate (FDR) estimation. A minimum of two unique peptides per protein were required for untargeted, whole-proteome quantitation, with a minimum peptide score (XCorr threshold) of at least 1.0.

For protein grouping, only peptide spectral matches with FDR ≤ 0.05 and a delta CN less than 0.15 were used. For quantitation, we included peptides with a coisolation interference of less than 75%.

For TIOX-enriched peptides, searching was performed as above, except that phosphorylation was included as a dynamic modification, with no minimum peptide per protein required. In addition, in Proteome Discoverer (v1.4.1.14), a phosphoRS search node was included to provide confidence of phosphosite localization.

**Normalization**

A table containing reporter ion intensities for all peptide matches was exported from Proteome Discoverer (tables S8 and S9) and parsed using custom scripts: A maximum coisolation filter of 75% was applied, and all values were scaled by scan injection time such that the reporter peak intensities were multiplied by scan injection time. For all reporter ion intensities passing an FDR filter of <0.05, the median reporter ion intensity for all peptides in each channel was used to normalize all channels toward the median and then apply a correction factor for each channel to all values in the table. Finally, peptides were filtered on whether the peptide was “used” by Proteome Discoverer. These values were then used to determine and generate a table for the summed reporter ion intensities for each protein grouping (tables S2 and S12). A summary of total peptides and protein groups identified in each experiment is summarized in table S13.

To determine how similar identical tissues from different animals were to one another, for each fish, the median summed reporter ion intensity for each protein group was determined, and then the log2 (tissue/median) value was taken for each fish. These log2-transformed ratios were then input into R for examination. First, density plots were generated of the log2 (tissue/median) values to examine their distribution centered around zero (fig. S2). Next, we clustered these values by tissue-specific abundance profile, computed the Euclidean distances, and performed complete-linkage hierarchical clustering. The resulting heat map (built with heatmap.2 function from gplots package, v2.17.0)
SUPPLEMENTARY MATERIALS

Generation of tissue-specific clusters

Unenriched, whole proteome. Because we were interested in examining the proteins of differential abundance among the three electric organs, the normalized log2 (tissue/median) values were further filtered such that only proteins with a minimum of twofold difference between the highest and lowest abundant electric organ were retained. These were input into R (v3.1.2) for k-means clustering and resulting heat map generation with the pheatmaps package (v1.0.2) (61), using Euclidean distance measurement and complete hierarchical clustering with k = 3. TiOX-enriched phosphoproteome. Using the normalized log2 (tissue/median) values generated above, we retained only the phosphorylated peptides that showed at least twofold difference between the highest-expressing and the lowest-expressing protein. An identical clustering method was used here, as was used for the unenriched, whole-proteome clustering, except that k = 4 was chosen.

Correlation of RNA expression and protein abundance values

To determine how the RNA expression values were obtained (see the “Calculation of RNA expression levels” section and table S5) compared to the protein abundance values obtained in these experiments, we gathered the RNA expression data for muscle, main electric organ, Sachs’ electric organ, and Hunter’s electric organ; we included RNA abundance estimates only in the gene models for which we also had protein abundance estimates (a total of 2866 proteins, once common contaminant hits were filtered out). We then took the log2 (electric organ/muscle) ratio for both the normalized RNA expression values and the normalized protein abundance values for both fishes [log2 (electric organ/muscle), starting with values in table S2]. To account for the difference in dynamic range between RNA sequencing and protein abundance data, a gene model was considered DE at the RNA level if it had a log2 (electric organ/muscle) value of greater than 2 or less than −2 (at least a fourfold difference); a protein was considered DE if it had a log2 (electric organ/muscle) ratio of greater than 1 or less than −1 (at least a twofold difference).

To visualize the correlation of the RNA and protein abundance data, the aforementioned criteria were used to classify each protein in each tissue (and in each fish). Five possibilities were considered for each gene model (protein group): (i) not DE at the RNA or protein level, (ii) DE at a protein level but not at the RNA level, (iii) DE at an RNA level but not at the protein level, (iv) DE at both the RNA and protein levels, and (v) DE at both the RNA and protein levels but opposite in direction. Using scripts, each gene model/protein was classified into each of these five bins, and a table of z values (for color specification) was created for input into R for building plots (Fig. 5).

REFERENCES AND NOTES


SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/7/e1700523/DC1

fig. S1. Phosphorylation sites in the C-terminal domain of E. electricus SCN4aa.
fig. S2. Normalization of proteomic and phosphoproteomic data.
table S1. Novel and known phosphosites in E. electricus proteins.
table S2. Median-normalized channel intensity values for unenriched, whole-proteome samples.
table S3. Normalized intensity ratios, log2 (tissue/median), on a per-peptide basis.
table S4. Potential roles for nonselectrogenic proteins and phosphopeptides differentially abundant in electric organs.
table S5. Expression values (RNA) for all predicted genes in assembly.
table S6. Intensity ratios, log2 (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.
table S8. Raw output from Proteome Discoverer, unenriched, whole-proteome samples.
table S9. Raw output from Proteome Discoverer, titanium dioxide enrichment, mass spectrometry, and phosphopeptide search.
table S10. Correlation of RNA and protein abundance values.
table S11. Comparison of new genome assembly and gene annotations to the previous assembly.
table S12. Median-normalized channel intensity values for titanium dioxide–enriched phosphoproteptides.

SCIENCE ADVANCES | RESEARCH ARTICLE


Downloaded from http://advances.sciencemag.org/ on November 17, 2017

Acknowledgments: We thank the University of Wisconsin-Madison Mass Spectrometry Facility for the aid in the study design and implementation and the University of Wisconsin-Madison Sequencing Facility for the efforts in generating the long-range paired-end 454 reads incorporated into the assembly in this study. Funding: This project has been funded by NSF grant MCB no. 1144012 (to M.R.S.), the Morgridge Graduate Fellowship (to L.L.T.), and the University of Wisconsin-Madison Genetics NIH Graduate Training grant (to L.L.T.). Author contributions: L.L.T. and M.R.S. conceived and designed the experiment. L.L.T. performed sample collection and preparation. G.A.B.-W. and M.R.S. wrote the manuscript. All authors edited the manuscript. M.R.S. supervised the project. Competing interests: The authors declare that they have no competing interests. Data and materials availability: The M5 proteomic and phosphoproteomic data have been deposited and are available publicly at the Chorus Project (www.chorussproject.org), under project ID #1325, titled “Electrophorus electricus quantitative proteomics and phosphoproteomics.” Raw sequencing reads have been uploaded to Sequence Read Archive (BioProject ID: PRJNA249073). The genome assembly and gene model predictions presented here are available at http://efishgenomics.zoology.msu.edu (Electrophorus electricus version 1.0).

Submitted 16 February 2017
Accepted 19 May 2017
Published 5 July 2017
10.1126/sciadv.1700523
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 and Michael R. Sussman

Sci Adv 3 (7), e1700523.
DOI: 10.1126/sciadv.1700523

ARTICLE TOOLS
http://advances.sciencemag.org/content/3/7/e1700523

SUPPLEMENTARY MATERIALS
http://advances.sciencemag.org/content/suppl/2017/06/29/3.7.e1700523.DC1

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions