Electric fish genomics: Progress, prospects, and new tools for neuroethology

William R. Pitchers *, Savvas J. Constantinou, Mauricio Losilla, Jason R. Gallant *

Dept. of Integrative Biology, Michigan State University, 288 Farm Lane RM 203, East Lansing, MI 48824, USA

ABSTRACT

Electric fish have served as a model system in biology since the 18th century, providing deep insight into the nature of bioelectrogenesis, the molecular structure of the synapse, and brain circuitry underlying complex behavior. Neuroethologists have collected extensive phenotypic data that span biological levels of analysis from molecules to ecosystems. This phenotypic data, together with genomic resources obtained over the past decades, have motivated new and exciting hypotheses that position the weakly electric fish model to address fundamental 21st century biological questions. This review article considers the molecular data collected for weakly electric fish over the past three decades, and the insights that data of this nature has motivated. For readers relatively new to molecular genetics techniques, we also provide a table of terminology aimed at clarifying the numerous acronyms and techniques that accompany this field. Next, we pose a research agenda for expanding genomic resources for electric fish research over the next 10 years. We conclude by considering some of the exciting research prospects for neuroethology that electric fish genomics may offer over the coming decades, if the electric fish community is successful in these endeavors.

Introduction

The powerful tools of genomics and genome manipulation are currently used nearly exclusively on a handful of well-established model species whose facile genetics and tractable husbandry allowed for the development of those tools. While this drives progress in those fields, the paucity of systems with reliable genomic data limits the insights that can be gained into the genetics of ecologically relevant traits, and may impede work on a number of central evolutionary problems. A core strength of the electric fish research, and of neuroethology as a discipline, is the focus on phenotype, broadly construed. This focus, combined with the strengths of electric fish as a model system, places neuroethology in a position to contribute to three of the five 'Grand Challenges' for biology as recently set out by the National Research Council (National Research Council, 2010) namely Connecting Genotype to Phenotype, Understanding the Brain, and Understanding Biological Diversity.

Since the discovery that weakly electric fish use electricity to sense their surroundings and communicate (Lissmann, 1958), researchers with interests and expertise spanning the range of biological disciplines have congregated around electric fish. Although the focus of contemporary work in the field has broadened considerably, electric fish researchers have continued to make valuable phenotypic and ecological connections that outpace many model organisms. The combination of genomics and the unique physiology of electric fishes – where the details of the electrosense link ecology and evolution intimately with neuroanatomy and ion channel kinetics (Fig. 1) – could allow for sweeping insights into how genotype connects to phenotype in an ecologically relevant system.

In the last few years, the availability of low-cost, high-throughput next-generation sequencing and sophisticated new molecular genetic techniques has laid the foundations for a ‘genomic renaissance’ in electric fish research. This paper will review the work of molecular biology in electric fish, beginning with the early biochemical contributions of Torpedo and Electrophorus and ending with whole genome sequencing efforts. In this work, the electric organ discharge (EOD) is a window into the neural system and the molecular workings of the electric organ (EO) and its component cells, and a view outward onto the ecology, behavior and evolution of the whole organism (Fig. 1). The latter part of the paper will then discuss the methods and benefits of integrating genomic and molecular tools into existing research programs.

1. Molecular biology of electric fishes: the first 30 years

1.1. The molecular biology of the neuromuscular junction

Torpedo rays (Miledi et al., 1971; and to a lesser extent Electrophorus electricus; Changeux et al., 1970) contributed deeply to our understanding of synaptic transduction. With their large EOs – Torpedo is “essentially a swimming purified acetylcholine receptor” (Miller, 2000) – these species provide researchers with the abundant source of receptor-rich membranes needed for describing the structure and biophysical properties of proteins at the neuromuscular junction (NMJ). Given the connection between EO and motor plates (Keesey, 2005), researchers could extract and purify the nicotinic acetylcholine receptor (nAChR) with relative ease (Sobel et al., 1977). These early biophysical studies were inherently comparative: immunohistological, ultrastructural (Rieger et al., 1976), and functional (Hess et al., 1982); however their efforts culminated in studies describing the detailed nanosstructure of the nAChR (Kistler and Stroud, 1981), visualization of the molecular subunits forming the ‘rosette’ around the ionophoretic channel (Kistler et al., 1982), and finally a full 3D model of the molecule (Mitra et al., 1989). In addition, this research lead to the characterization of a number of other important proteins at the NMJ – notably agrin (Nitkin et al., 1987), dynein (Mou et al., 1998), and rapsyn (Elliott et al., 1980).

Beginning in 2007, Nazarian et al. (2011, 2007) demonstrated the potential for using genomic tools to probe the identity of NMJ proteins using high throughput Sanger sequencing. Nazarian et al. (2011) built a database of NMJ-associated proteins and transcripts present in electroplaques of T. californica, enabling the characterization of a suite of mammalian expressed sequence tags, and proteins of unknown function, as being associated with the NMJ. These discoveries proved crucial to the fields of biophysics and biomedicine (reviewed by Changeux (2012)), but although this work is indirectly relevant to neuroethology, it is the weakly electric fishes that have received the majority of the attention from the field.

1.2. Molecular biology of the EO: effectors and modulators

EOD duration varies 100-fold among species (Hopkins, 1999). This diversity is due in large part to variation in the electrocyclic Na⁺ currents (Ferrari et al., 1995), which in turn are regulated by voltage-gated Na⁺ channels, and these are among the best studied electrocyte proteins. Research using polymerase chain reaction (PCR), quantitative PCR (qPCR), cloning, and first-generation sequencing has provided sequences of Na⁺ channel genes from around 20 species of weakly electric fishes, representing both Gymnotiformes and Mormyroids (Arnegard et al., 2010; Lopreato et al., 2001; Zakon et al., 2006). These studies provide a striking example of parallel molecular evolution: independent neofunctionalization (Ohno, 2013) of the voltage-gated Na⁺ channel gene scn4aa in both lineages.

The teleost-specific whole-genome duplication event (Section 1.4) left these taxa with two paralogs (scn4aa and scn4ab) of the tetrapod muscle Na⁺ channel gene Na1.4 (Lopreato et al., 2001; Novak et al., 2006). Following down-regulation of scn4aa in the ancestors of both lineages (Thompson et al., 2014), the expression of this paralog was found to be restricted to the EO (Zakon et al., 2006). Freed from the selective constraint to maintain the muscle phenotype (as scn4ab retained this function), scn4aa experienced strong positive selection on amino acid replacements critical for Na⁺ channel kinetics, allowing for the evolution of EOD waveform diversification and signal complexity in both lineages (Arnegard et al., 2010; Zakon et al., 2008).

This subfunctionalization (evolutionary repurposing of duplicate genes; Magadum et al., 2013; Ohno, 2013) of membrane ion channels allows for greater EOD complexity and faster firing rates, but these properties are not fixed, even within a species. The EOD frequently varies between sexes and seasons, and is regulated by hormones (Bass and Hopkins, 1984; Hopkins, 1972). Furthermore, many studies have demonstrated that a variety of hormones have effects on the EOD (Allee et al., 2008; Bastian et al., 2001; Deemyad et al., 2013; Dulka et al., 1995; Dunlap et al., 2006; Dunlap and Zakon, 1998; Maler and Ellis, 1987; Markham et al., 2009a; Mills et al., 2013; Dulka et al., 1995; Dunlap and Zakon, 2006; Dunlap and Zakon, 1998; Maler and Ellis, 1987; Markham et al., 2009a; Mills and Zakon, 1987; Smith and Combs, 2008; Telgkamp et al., 2007; ZuPanc, 2002).

Changes to the EOD mediated by hormones have been recorded over a period of minutes (Markham and Stoddard, 2005), days or weeks (Dunlap et al., 1997; Ferrari et al., 1995; McAnelly and Zakon, 2007), and each is regulated differently. Changes occurring on the timescale of minutes are regulated by trafficking Na⁺ channels into the electrocyte membrane (Markham et al., 2009b), and is circadian and socially controlled in Sternopygus macrurus. Circadian variation in EODs, at least in Brachycephalopus spp. (Franchina and Stoddard, 1998; Stoddard et al., 2007), is probably mediated by glutamate (Silva et al., 2008) and vasotocin (Perrone, 2010). Short-term hormonal EOD modulation seems to be mediated through cAMP and PKA after G-coupled receptor activation (Markham and Stoddard, 2005; McAnelly et al., 2003; McAnelly and Zakon, 1996).

Sexual dimorphism in EODs occurs in both Gymnotiformes (Allee et al., 2009; Ho et al., 2010; Markham and Stoddard, 2013; Smith, 2013) and mormyrids (Bass and Hopkins, 1985), and while androgens appear to be involved, other factors are implicated in mediating the dimorphism (Allee et al., 2009). Sexual dimorphism appears to have diverged rapidly, at least among Apterontus spp. (Ho et al., 2010). The EOD (Carlson et al., 2000; Ho et al., 2010)
and its dimorphism (Franchina et al., 2001; Terleph, 2003) are also plastic in response to social cues, and this too appears to be hormone-mediated, with cortisol implicated (at least in Brachyhypopomus; Salazar and Stoddard, 2009). We refer readers to the excellent review by Markham (2013) for a more detailed exploration of this topic.

1.3. Evolution and development of sensory systems

Understanding electrolocation and electrocommunication requires studying not only EOs, but also the electrosensory organs and the nervous system that serves them. To this end, numerous studies have used in situ hybridization or immunolabelling to localize a range of crucial molecules, including neurotransmitters and neurotransmitter receptors (Bottai et al., 1997; Johnston et al., 1990; Maler and Monaghan, 1991; Sas et al., 1990; Stroh and Zupanc, 1995); ion channels (Engelmann et al., 2008; Rashid et al., 2001); Ca²⁺ binding proteins (Friedman and Kawasaki, 1997; Smith et al., 2000) and neuropeptides (Sas et al., 1990; Stroh and Zupanc, 1995; Yamamoto et al., 1992; Zupanc et al., 1991). Zupanc (2002) reports more than fifteen neurotransmitter substances in the central posterior/pre-pacemaker nucleus (CP/PPh) of gymnotiform fish of the genera Apteronotus and Eigenmannia. The central nervous system transcriptome of Apteronotus leptorhynchus has been assembled and validated by shotgun proteomics (Salisbury et al., 2015), thus producing a resource that may aid in exploring the underlying cellular mechanisms of critical neurobiological processes, such as adult neurogenesis, neuronal regeneration, and the neural basis of behavior. The origin of brain structures devoted to processing electrosensory information currently remain unknown (Carlson et al., 2011).

Electrosensitivity requires specialized sensory cells, and weakly electric fish have both high and low frequency tuned receptors (Gibbs, 2004; Jørgensen, 2005). In addition to bearing multiple types of receptors as an adult, there is evidence that the form and specificity of electrosensory receptors vary over an individual fish’s growth and metamorphosis (Bensouilah et al., 2002; Denizot et al., 2007; Painter and Kramer, 2003; Szabo, 1965). Given the apparent complexity of these systems, there is surprisingly little molecular data on electric fish electrosensory receptor development or the mechanism of electrosensory signal transduction (unknown at the time of writing). Some expression data for RNA and protein in the electrosensory cells of the weakly electric skate Leucoraja erinacea has been collected (Baker et al., 2013; Gillis et al., 2012; Maxwell et al., 2008), but there is more to be learned with transcriptomic/molecular techniques.

1.4. Evolution and development of novel electrogenic systems

Most EOs develop from myoblasts, i.e., skeletal muscle precursors (Kirschbaum, 1977; Szabo, 1960). As such, electrocytes share similarities with myocytes: they are multinucleated, excitable, and retain several muscular proteins. However, electrocytes are larger, non-contractile, more highly-polarized and more variable morphologically. How EOs have evolved from, and differ molecularly from skeletal muscle is a central question for electric fish researchers. EOs evolved independently in mormyrids and Gymnotiformes (Fig. 2) following the teleost-specific whole genome duplication event (Amores et al., 1998; Hoegg et al., 2004; Hurley et al., 2007; Lopreato et al., 2001; Taylor et al., 2001). Studying the expression of the resulting paralogs across a range of species (e.g., Thompson et al., 2014) addresses the importance of sub-/neo-functionalization as a source of evolutionary novelty (Magadum et al., 2013; Ohno, 2013; Zhang, 2003).

The two best-studied aspects of EO differentiation are the dependence on innervation and the suppression of muscle-specific protein expression. Sternopygus macrurus requires EO innervation for its differentiation and maintenance; denervating the EO causes the appearance of sarcomeric proteins in mature electrocytes, and inhibits differentiation of new electrocytes from satellite cells (Patterson and Zakon, 1993; Unguez and Zakon, 2002, 1998a; Weber et al., 2012; Zakon and Unguez, 1999). By contrast, the electrocytes of the Mormyrid Pollimyrus isidori do not require innervation for differentiation (Szabo and Kirschbaum, 1983), and histological observations on eight Gymnotiformes show that electrocytes differentiate before they are innervated (Kirschbaum and Schwassmann, 2008; Schwassmann et al., 2014).

Researchers have documented the mechanism of muscular protein suppression in S. macrurus. Initial immunolabelling and ultrastructural studies demonstrated that electrocytes express several muscle proteins, but lack some sarcomeric proteins, sarcomeres and T-tubules (Cuellar et al., 2006; Kim et al., 2004, 2008; Patterson and Zakon, 1996; Unguez and Zakon, 1998a, 1998b). Nevertheless, electrocytes transcribe the genes for the muscular proteins that they lack (Cuellar et al., 2006). Moreover, the myogenic regulatory factor genes, which control skeletal muscle gene expression, are all transcribed in the EO (Kim et al., 2008, 2004). The electrocyte transcript and protein profiles do not match, which suggests post-transcriptional mechanisms are responsible for the EO differentiation from muscle in S. macrurus.

Pinch et al. (2016) used genomic tools (both qRT-PCR and RNA-seq) to examine the regulation of sarcomere gene expression in S. macrurus by comparing EO and muscle transcription profiles of miRNA and mRNA. They did not detect differential expression in sarcomere-related genes between the two tissues, confirming earlier work in this species. In addition, protein-degradation pathways showed similar expression in both tissues, which implies that post-translational sarcomeric protein degradation does not regulate the electrocyte phenotype. They also identified 155 annotated miRNAs, all of which were expressed in both tissues. Three miRNAs were upregulated in the EO, and they were known to participate in muscle differentiation and the down-regulation of sarcomere-associated genes. The research on the non-contractile phenotype of electrocytes in S. macrurus unambiguously points toward a post-transcriptional mechanism of regulation via silencing a few sarcomeric genes.

While most data on EO developmental mechanisms has been collected in S. macrurus, these mechanisms may not be representative of all electric fish. For instance, Briemiaurus brachyistius expresses both transcripts and proteins from some muscle-specific genes while decreasing the expression of others (Gallant et al., 2012, 2014), suggesting that a transcriptional regulatory mechanism downregulates the expression of some muscle genes. Work on the mormyrids Campylomormyrus compressirostris and Campylomormyrus tshokwe (Lamanna et al., 2015), on the Gymnotiformes E. electricus and Eigenmannia virescens, and in the Siluriform Malapterurus electricus (Gallant et al., 2014), suggests that they share a similar transcriptional regulatory mechanism to that of B. brachyistius.

1.5. Unraveling the phylogeny of weakly electric fish

Electric fish can be problematic for taxonomists, with morphologically cryptic species not uncommon (Feulner et al., 2006; Nagamachi et al., 2010; Silva et al., 2015) and many phylogenetic details unresolved. To help resolve the relationships among species and populations, researchers have used genetic markers to infer relatedness. Microsatellites (or simple sequence repeats, SSRs) are short, repeated sequence motifs. These repetitions tend to accumulate mutations comparatively quickly and are therefore most used in situations where the relationships among subjects are expected to be close, such as assessing divergence among populations
Arnegard et al., 2005; Feulner et al., 2006, 2008). Amplified fragment length polymorphisms (AFLPs) are neutral markers used to detect polymorphisms among more distant relationships. DNA is fragmented using restriction enzymes, and the pattern of resulting fragments depends on the location of the restriction sites where the enzymes bind. Since these restriction sites evolve more slowly than SSRs, they are often used to infer relationships among species (Lavoué et al., 2008; Sullivan et al., 2004).

At larger scales, such as among genera, the sequence of the mitochondrial protein cytochrome b (Sullivan et al., 2004, 2000) and other mitochondrial sequences (Alves-Gomes and Hopkins, 1997; Lamanna et al., 2014; Lavoué et al., 2012) are used, since these parts of the genome evolve slower still. Molecular markers can also be used in combination with other potentially informative data such as the analysis of EOD recordings (Alves-Gomes and Hopkins, 1997; Feulner et al., 2008, 2006; Sullivan et al., 2000), geometric morphometrics (Feulner et al., 2008), and even fossil data (Lavoué et al., 2012).

Marker-assisted phylogenetic studies have sometimes supported existing species identifications based on morphology (Feulner et al., 2008; Sullivan et al., 2000) and have unmasked cryptic morphs elsewhere (Feulner et al., 2006). These studies also have the potential to reveal surprises, such as mitochondrial introgression between putative species (Lavoué et al., 2008; Sullivan et al., 2004) and a pattern of reduced variability among the EOs of Petrocephalines when compared to the Mormyrines. The potential for NGS technology to contribute to phylogenetic studies is demonstrated in a transcriptomic study where the complete mitochondrial genome for C. compressirostris was assembled and >1500 SSR markers were developed (Lamanna et al., 2014). With this density of markers, studies can effectively probe the structure of populations and search for divergent regions or loci that show a signature of selection (Ellegren, 2014).

1.6. Macroevolution & diversification

The studies above demonstrate the integration between levels of analysis – molecular, cellular, and at the level of tissues/organs – that form the ‘integrated phenotype’ (Fig. 1). This integration is a key strength of weakly electric fish as a research system, but it can be extended; NGS techniques greatly facilitate the use of electric fish to address broad genetic and evolutionary questions. The ‘natural experiment’ in repeated independent evolution of the EO is well-suited for comparative molecular work, since the parallel evolution of Gymnotiformes and Mormyroids allows for the separation of phylogenetic ‘noise’ from the evolutionary ‘signal’.

Gallant et al. (2014) took this approach in examining the genomic basis of anatomical and physiological convergence across three independently evolved electric fish lineages. The authors assembled a draft version of the E. electricus genome and EO and skeletal muscle transcriptomes from three Gymnotiformes, one mormyrid and one Siluriform (see Section 1.4). Comparative analyses revealed genes potentially involved in the phenotypic characteristics of electrocytes that differentiate them from muscle, and suggested that a common regulatory network of transcription factors and developmental pathways may have been repeatedly utilized in the evolution of EOs (Gallant et al., 2014).

Work on the Electrophorus electricus genome (Gallant et al., 2014; Traeger et al., 2015) has also produced the first analysis of electric fish miRNA expression (Traeger et al., 2015). Researchers examined transcriptomes from eight adult tissues, including all three EOs and detected upregulated genes in these EOs that show a signature of selection (Ellegren, 2014).

![Fig. 2. The distribution of available genomic data for electric fish. Genomes, transcriptomes, population genetics marker sets, partial sequences and proteomic datasets - 'X' indicates a single dataset, 'XX' indicates multiple datasets.](http://dx.doi.org/10.1016/j.jphysparis.2016.10.003)
enriched in functions related to transmembrane transport, androgen binding, and receptor signaling. They also described a unique collection of expressed miRNA in each of the three EOs. This included three conserved miRNAs known to participate in the inhibition of mammalian muscle development, and one highly expressed novel miRNA.

2. Developing a ‘Genomic Toolkit’ for electric fish: a research agenda for the next decade

The integrated phenotype concept in electric fish (Fig. 1) makes them a powerful model system for addressing research challenges across biological disciplines, and more specifically in neuroethological contexts. Genomic and molecular data have accelerated in the past 5 years, largely due to the rise of low-cost next generation sequencing (NGS) approaches. The combination of the strengths of the electric fish model with coordinated, strategic advances in both data collection and protocol development, would greatly facilitate the ability to address 21st century research objectives (National Research Council, 2010), particularly with regard to understanding the brain, and integrating geno-phenotype with phenotype across levels of biological analysis. This coordinated research effort can be imagined as a ‘genomics toolkit’ consisting of established protocols to understand sequence divergence, assess and manipulate gene location/expression, and generate transgenic tools. The success of these endeavors over the next decade will undoubtedly usher in a new ‘golden age’ for electric fish biology. Such a genomic toolkit is only as useful as it is accessible, highlighting a need for community involvement and support for new standards for data sharing and protocol exchange.

2.1. Additional sequencing

The growth and coordination of sequence information are crucial for bringing electric fish biology into the new ‘golden age’ promised by genomic tools. At this point, genomic sequencing should focus on species that can be bred in captivity (for production of transgenic lines), that are readily available, and that represent the principal taxonomic groups. Sequencing the genomes of fish with these characteristics will allow us to refine genetic tools, establish detailed protocols, and prime research for other electric fish species.

With NGS technologies becoming increasingly affordable, even small labs can acquire vast amounts of sequence data. Vertebrate genomes sequenced with 50x coverage can return 100s of gigabytes of data. Further, in the process of, e.g., assembling a genome draft from this data, the space required for the various intermediate files can easily grow to more than a terabyte. Properly planning an NGS project therefore also involves planning for data management and backup. Many Universities maintain computing facilities and/or are connected to local or national grids, and we would therefore encourage researchers to acquaint themselves with these formats have in common is that data is stored as plain text, with genomic data should familiarize themselves with these. What these formats have in common is that data is stored as plain text, files that are often too large to be opened on a personal computer. Working with data in these formats (and most other genomics formats) therefore principally involves programmatically manipulating files without opening them; for which the Unix environment is well-suited. Rather than recommending specific software packages, we would therefore encourage researchers to acquaint themselves with the Unix ‘shell’, and with at least one of the most commonly-used programming languages, such as Python, Perl or Ruby. For newcomers to genomics, a general computing skills text would be a good place to start (Haddock and Dunn, 2011 Buffalo, 2015 are highly recommended), and a plurality of working bioinformatics share tools and advice at websites like Biostar, SEQanswers and Stack Overflow.

2.2. Molecular biology applications post-sequencing

Sequence information has multiple uses. It can be used to generate PCR/qPCR primers for population genetic marker development (see Section 1.5), in situ hybridization templates, for differential expression and RNA profiling, comparative genomics, and identification of targets for loss of function studies and generation of transgenic lines.

Using specific gene sequence information, primers can be designed to amplify the focal gene specifically through PCR/qPCR. Identification of high quality PCR primers is necessary, due to the use for PCR in generating templates (e.g., for in situ probes) and assaying experimental outcomes. PCR can identify NGS-predicted alternative splice variants and be utilized to assay potential single nucleotide polymorphisms (SNPs). qPCR allows the quantification of DNA in real time showcasing its use as an assay of gene expression; however, qPCR can also be used for validating microarray experiments and monitoring biomarkers (VanGuilder et al., 2008).

With RNAseq, researchers can discover and quantify the full set of transcripts – including alternatively spliced isoforms and non-coding RNAs – from a sample (Martin and Wang, 2011). One common use for RNAseq data is differential expression analysis, wherein the expression level of every gene is compared statistically across two or more groups of samples (Mortazavi et al.,...
A brief glossary of genomics-related terms, and recommended reading for those wishing to learn more.

<table>
<thead>
<tr>
<th>Genomics</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-generation sequencing</td>
<td>&quot;Sanger sequencing&quot;; a long read technology still frequently used to sequence plasmids, PCR products and other relatively short DNAs</td>
</tr>
<tr>
<td>Next-generation sequencing (NGS)</td>
<td>A suite of technologies that use amplified sample DNA/RNA as a template (no need for sample cloning) and run millions of sequencing reactions in parallel. Sequence is produced in fragments</td>
</tr>
<tr>
<td>Library</td>
<td>A set of templates produced by shearing sample DNA into fragments and ligating oligo adapters at their free ends</td>
</tr>
<tr>
<td>Template</td>
<td>Recombinant DNA molecule made up of an adaptor sequence followed by the target sequence (typically an unknown fragment)</td>
</tr>
<tr>
<td>Adapters</td>
<td>Synthetic leading/trailing sequences added to fragments to facilitate sequencing chemistry</td>
</tr>
<tr>
<td>Barcodes reads</td>
<td>The individual string of base-pair codes that are generated in parallel by NGS technologies. Require trimming and alignment before analysis</td>
</tr>
<tr>
<td>.fastq</td>
<td>The file format output by the vast majority of sequencing facilities. A plain text file that includes both base calls and a quality score for each base</td>
</tr>
<tr>
<td>.fasta</td>
<td>A plain text file format for containing named/labeled DNA or RNA sequences</td>
</tr>
<tr>
<td>Mapping/Alignment</td>
<td>The process of aligning short reads to a reference sequence, which could be a complete genome, transcriptome, or de novo assembly</td>
</tr>
<tr>
<td>De novo assembly</td>
<td>The computational rebuilding of the genome/transcriptome from the sequenced fragments</td>
</tr>
<tr>
<td>Assemblers</td>
<td>Softwares used to de novo assemble fragments together - many tools available with various strengths/weaknesses</td>
</tr>
<tr>
<td>.sam/.bam</td>
<td>SAM is a format for storing sequence data in a series of tab delimited text columns. Currently the most common output format from mappers/aligners. BAM is a binary version: non-human-readable but much smaller files</td>
</tr>
<tr>
<td>.vcf</td>
<td>A plain text file format that lists only variant sites (rows) by a list of samples (columns) Indexing of genomic and transcriptomic features by base-pair coordinates on raw genome or transcript sequence. Often formatted in .gff or .gtf format</td>
</tr>
<tr>
<td>.annotation</td>
<td>Everything transcribed from the genome; primarily mRNA and rRNA, but containing many classes of small and large non-coding RNAs</td>
</tr>
<tr>
<td>Gene ontology</td>
<td>‘GO’ annotations: a way of assigning or analyzing functional gene annotations by comparing them to known biological pathways</td>
</tr>
<tr>
<td>transcriptome</td>
<td>Sequencing technologies by Illumina, and currently the standard at most sequencing facilities. Results in relatively short reads (up to 200 bp), with a lower error rate than PacBio.</td>
</tr>
<tr>
<td>‘HiSeq’ or ‘miSeq’</td>
<td>Sequencing technologies from ThermoFisher/Life Technologies. Similar in read length to Illumina (max. 400 bp), and often slightly more affordable, but is subject to increased errors in homopolymer runs.</td>
</tr>
<tr>
<td>Ion Torrent</td>
<td>Sequencing technologies from Illumina. Generally shorter reads (up to 200 bp), with a much higher error rate than Illumina.</td>
</tr>
<tr>
<td>'PacBio'</td>
<td>Relatively new sequencing technology from Pacific Biosciences, that features very long reads (average 10,000 bp), but a much higher error rate than Illumina.</td>
</tr>
<tr>
<td>RADseq</td>
<td>Restriction Associated DNAseq. By digesting genomic DNA with a specific enzyme before sequencing, a subset of the genome is sequenced with higher coverage. Useful for ascertaining intraspecies variation</td>
</tr>
<tr>
<td>TagSeq</td>
<td>The RNAseq technique of sequencing only the 3' end of miRNAs, such that the overall amount of sequence per individual is reduced, but the number of fragments sequenced is maintained. However, the ability to detect splice variants and some other features is lost</td>
</tr>
</tbody>
</table>

**Recommended reading**

- Bild et al., 2014; Borgman et al., 2012; Bradnam et al., 2013; Davey and Blaxter, 2011; De Wit et al., 2015; Earl et al., 2011; Ellegren, 2014; Ekblom and Galindo, 2010; Ekblom and Wolf, 2014; Field et al., 2009; Fonseca et al., 2012; Hatem et al., 2013; Kaye et al., 2009; Lee et al., 2016; Liu, 2005; Lohman et al., 2016; Macaulay and Voet, 2014; Martin and Jiggins, 2013; Metzker, 2009; Narum et al., 2013; Sims et al., 2014; Wake, 2008; Whitlock et al., 2010

**Gene ontology**

- **GO** annotations: a way of assigning or analyzing functional gene annotations by comparing them to known biological pathways
- **transcriptome** Everything transcribed from the genome; primarily mRNA and rRNA, but containing many classes of small and large non-coding RNAs
- **Gene ontology** ‘GO’ annotations: a way of assigning or analyzing functional gene annotations by comparing them to known biological pathways

**Wet-Lab Molecular Biology**

<table>
<thead>
<tr>
<th>Genetic markers</th>
<th>Microsatellites, Simple sequence repeats, or AFLPs that can be easily assayed and used as a proxy or tag for the genotype of regions surrounding it</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral markers</td>
<td>DNA markers in non-coding and presumed non-regulatory regions</td>
</tr>
<tr>
<td>Comparative genomics</td>
<td>Comparing the structure, organization and gene content of genomes in a tree-wise fashion to elucidate trends and patterns of genome evolution</td>
</tr>
<tr>
<td>Splice variants</td>
<td>Multiple mature mRNAs derived from the same precursor RNA</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA: a very short RNA molecule (22 bp) involved in RNA silencing</td>
</tr>
<tr>
<td>RNAseq</td>
<td>Next-gen sequencing where RNA (transcriptome) is sequenced rather than DNA (genome). May be further refined to target specific families and types of RNAs</td>
</tr>
<tr>
<td>In situ hybridization</td>
<td>Allowing labeled DNA or RNA probes to anneal to DNA/RNA targets in vivo. Frequently used to localize mRNA transcripts to tissues</td>
</tr>
<tr>
<td>antisense reporter</td>
<td>Complementary to an mRNA sequence. Double stranded mRNA is not effectively translated, and is targeted by the cell for degradation</td>
</tr>
</tbody>
</table>

**Recommended reading**

- Deisseroth, 2010; Deisseroth et al., 2006; Heidenreich and Zhang, 2015; Mancuso et al., 2010

This allows for the association of the differentially expressed genes with the differences between any two groups; e.g., different tissues, species, developmental stages, diseased or transgenic individuals, cell activation states. However, the identification of splice variants may also prove useful to neuroethologists in light of the emerging role of alternative splicing in neural patterning (reviewed by Li et al., 2007). The availability of genomic sequences for electric fish allows for the use of comparative genomics analyses. For instance, regions of the genome that are highly conserved between divergent populations likely correlate with shared phenotypes. Selective events leave molecular signatures, (e.g., regions of reduced genetic diversity) that can be used to identify genes controlling EOD differences or other traits.

Further to these uses, sequence information is a prerequisite for planning follow-up molecular studies, such as loss of function manipulations (Section 2.3) and generating transgenic animals (Section 2.4).
2.3. Develop resources for transient loss of function studies

Loss of function studies enable researchers to assign a potential function to a gene by assaying how the cell/tissue/organism responds in the gene’s absence. Transient loss of function studies mediate gene knockdowns via exogenously injected oligomers, allowing for transient effects on gene function. Tracking other gene/protein expression changes after transient loss of function studies can determine gene regulatory networks, highlighting again the remarkable capacity for large-scale biological processes (e.g., communication, sexual selection, neural coding) to be understood at the gene level in electric fish (Fig. 1). The most widely used oligomers that are injected to generate loss of function are morpholinos and double stranded RNA (dsRNA).

Phosphorodiimidate morpholino oligomers (or simply ‘morpholinos’) are small (~25 bp) oligomers that bind to specific RNA sequences, thereby interrupting post-transcriptional processing using an antisense technology (see review by Moulton (2016)). Knockdown of the targeted gene product can be achieved by blocking the translation of the targeted RNA sequence or by blocking the spliceosome and causing intron retention (Fig. 3). Morpholinos can be injected into single cell/developing embryos or adult tissues, without permanent effect. ‘Scrambled’ morpholinos are available to purchase for use as a control for morpholino-specific effects.

Vivo-morpholinos are synthesized with a carrier subunit that allows for delivery to target cells after intramuscular, intravascular, or intraperitoneal injections (Kotb et al., 2016; Notch et al., 2011; Schulman et al., 2016) and even into developing fish embryos with a bath immersion (Wong and Zohar, 2015). Modifications to the morpholino subunits can generate a tool for inducible gene knockdowns (Shestopalov et al., 2012), providing extra specificity and control; e.g., fluorescein-tagged morpholinos can be visualized as well as electroporated (Hyde et al., 2012). Correctly identifying targets for morpholino loss of function requires access to sequence data; either the coding sequence (translation-blocking morpholinos) or an annotated genome (to identify splice-blocking targets).

Like morpholinos, RNA interference (RNAi) is a type of antisense technology used to knockdown gene expression. RNAi consists of injecting long double stranded RNA (dsRNA) or short dsRNAs (siRNAs, miRNAs, piRNAs) into an organism/cell. When the RNAi pathway is activated (Fig. 3), the dsRNA is cut by the enzyme DICER into small pieces, essentially generating many siRNAs that target the entire dsRNA sequence after forming the RNA-induced silencing complex (RISC) with other proteins (Alagia and Eritja, 2016; Kim et al., 2009). RNAi components have been manufactured that permit more precise spatiotemporal control and have been utilized in zebrafish (Andrews et al., 2014; Tang et al., 2007).

RNAi loss of function is similar to that of morpholinos; the effects are transient and affect protein expression through RNA, and thus could be used on species that only have transcriptomic data available. RNAi may be preferable to morpholinos due to lower costs and may be more viable for larger screens; e.g., De Rienzo et al. (2012) championed their use in zebrafish as being rapid in effect, inexpensive and a tissue-specific method beyond embryonic stages. However, RNA can be difficult to work with (as it is less stable than DNA) and cloning the sequence of interest is required.

One of the strengths of transient loss of function is the ability to induce loss of function in adult organisms. Because many electric fishes are difficult to breed in captivity, morpholinos and RNAi will likely be popular tools – and could easily be utilized in the field. These tools could be applied to understand how gene regulated changes to the EOD affect sexual selection, electrolocation, and electrocommunication.

2.4. Develop resources for manipulating the genome

While some studies require transient/post-developmental gene manipulation, other studies depend on constitutive gene expression or mutant lines. CRISPR/Cas9 (or simply CRISPR) enables researchers to generate transgenic lines and mutants. CRISPR originated as a prokaryotic immune system complex used as defense from foreign DNA (Barrangou et al., 2007), but has been modified for laboratory use. A single guide RNA (sgRNA) recognizes and binds with a specific sequence on the DNA along with a Cas9 endonuclease. This complex, through Cas9, then introduces double

---

Fig. 3. Schematic summary of morpholino, RNAi, and CRISPR approaches for gene/genome function manipulation. See text (Sections 2.3 and 2.4) for a description of these approaches. 1 Knockdown utilizing morpholinos and RNAi can also occur in adult tissues overtime (Section 2.3). 2 CRISPR can also be performed via in vitro transcription of just sgRNA and subsequent injection of the sgRNA/Cas9 enzyme complex.

stranded breaks at that precise location. The double stranded break is repaired through the nonhomologous end-joining (NHEJ) pathway, ultimately rejoining the DNA but often introducing insertions, deletions, or both (indels), or, if a homologous donor sequence is present at the time of repair, the homology-directed repair (HDR) pathway can be activated and the genome may insert the donor sequence (Fig. 3; see review by Sander and Joung (2014)). Through NHEJ repair, mutant lines (knockouts) can be generated and HDR allows for transgenic lines to be produced.

One strength of CRISPR is that, unlike mutagenesis screens, specific genome regions can be targeted, and a range of knockout/knock-in can be made by modifying the genomic sequence in vivo (Mei et al., 2016; Ran et al., 2013; Tsai and Joung, 2016). Genomic edits are passed on to all the daughters of the treated cells; if incorporated into the gametocytes, genomic edits can be transmitted to all progeny, producing breeding lines with stable knockout and knock-ins (Auer et al., 2014; Hwang et al., 2013; Kimura et al., 2014; Li et al., 2015). Knockouts through CRISPR can be highly efficient (75–99%) even for multiple genes in a single embryo (Jao et al., 2013). The effects are often biallelic, disrupting function completely and almost immediately (necessary for developmental studies). Current advances in CRISPR protocols have generated an inexpensive, cloning-free method for generating sgRNA and Cas9 mRNA/protein that most labs can readily adopt (Gagnon et al., 2014; see supplement for detailed protocol). CRISPR modifications occur in the genome, thus genomic sequence data is strongly recommended for identifying high quality CRISPR targets.

The potential applications for CRISPR have been expanded by recent molecular innovations (Harrison et al., 2014; Hsu et al., 2014). For example, using CRISPR it could be possible to visualize protein expression in vivo by inserting reporters, offering more information in concert with gene expression visualized by in situ hybridization. Green fluorescent protein (GFP) tagged proteins, coexpression of GFP with a gene of interest, or GFP under control of an endogenous gene promoter can be visualized with a fluorescent microscope to determine location of the gene of protein transcript. Comparing both gene transcript and protein location can uncover surprising results (see Section 1.4); localizing both gene outputs in a single tissue would be much easier, and more informative, by performing in situ hybridization on a transgenic fish that always expresses fluorescent tagged protein for the gene of interest.

Not only can CRISPR facilitate protein localization through reporter knock-ins, but it can also be used to identify the sequence basis of protein properties. Zakon et al. (2008) note that we currently do not know how the positively selected substitutions on the electrocyte-expressed scn4aa alter the biophysical properties of the Na+ currents. CRISPR site-directed mutagenesis could identify the effect of single nucleotide polymorphisms on ion channel kinetics: either in vivo, or in vitro through exogenous expression (e.g., in a Xenopus oocyte). Such studies may have direct consequences beyond neuroethological and electric fish research, because several of these amino acid replacements occur near sites associated with disease-causing mutations in human Na+ channel genes (Zakon et al., 2006).

Similarly, CRISPR can replace not only SNPs, but entire gene sequences. Candidate genes for species divergence could be reciprocally inserted for definitive gene function assays, or to test for genetic effects of different genetic backgrounds. Linking changes in protein sequence to EOD signaling changes could facilitate studies exploring sexual selection and EOD evolution among species; e.g., by driving male-like expression of candidate genes in female fish and vice versa. Finally, sgRNA’s can target nearly any location in the genome, opening the possibility for studies on differences in non-coding/cis-regulatory regions.

2.5. Limitations of morpholinos, RNAi, and CRISPR

Transient loss of function tools and genomic editing are powerful; however, like any tool, each has its own limitations. Below we present several limitations, but researchers are encouraged to read the current literature on these issues while deciding which tool(s) to use.

Reports that compare morpholino efficiency to CRISPR or mutant strains (Kok et al., 2015) suggest there are potential off-target effects and phenotypes when using morpholino (but see Morcos et al., 2015). Like morpholinos, there are known off-target effects seen with RNAi (Jackson et al., 2003; Nunes et al., 2013); therefore, proper controls and careful interpretation of results are absolutely necessary.

The availability of single cell embryos is a necessity for CRISPR. Electric fishes of various species have been bred successfully, including multiple mormyrids and Gymnotiformes; Kirschbaum and Schugardt (2002) provide a relatively comprehensive list of all laboratory bred electric fish. Breeding of these fishes is not without difficulty; most freshwater electric fish breed during the rainy season and it is necessary to simulate some of these conditions to induce gonadal recrudescence. At a minimum, the water conductivity must be dropped (Schugardt and Kirschbaum, 2004), although some Gymnotiform species seem to also require an increase in water depth (Kirschbaum and Schugardt, 2002). Some effort has been made to generate developmental series for electric fishes (Diedhiou et al., 2007), allowing more rigorous developmental hypotheses to be tested in a systematic way among labs.

CRISPR has historical issues of non-specificity and off-target effects (Zhang et al., 2015). These have been addressed in part through the double-nickase approach (Mei et al., 2016; Ran et al., 2013; Tsai and Joung, 2016); however, this approach limits available targets. The knockdown efficiency of CRISPR can vary among target regions (Doench et al., 2014; Hua Fu et al., 2014), with some target genes currently inaccessible to genomic edits. Without developing inducible CRISPR constructs, embryonic-lethal genes cannot be thoroughly studied with knockouts, but titrations of dsRNA or morpholino could be used in such cases.

2.6. Community dissemination

Developing genomic tools for electric fish benefits the electric fish community as a whole and requires collaboration among researchers. Compiling and curating available electric fish sequence data will likely maximize the value of the data to the community (Field et al., 2009; Kaye et al., 2009; Whitlock et al., 2010), and encourage comparative studies as seen in other model organism communities such as ZFIN ( Howe et al., 2012), WORMBOOK (Greenwald, 2016), and FLYBASE (Attrill et al., 2016).

With an annotated genome and multiple assembled and annotated transcriptomes now available for electric fishes (and more currently being assembled – Gallant; unpublished data), the computational development of markers for future phylogenetic studies should be much more efficient. A collection of electric fish nucleotide and protein databases are available at http://efishgenomics.integreativelybiology.msu.edu/ and submissions of sequence data are welcomed (visit the site for details).

To maximize the value of resources, researchers need an efficient way to share genomic tools and established protocols. Like the aforementioned databases for other organisms, a putative ‘EFISHBASE’ should include databases and repositories for transgenic lines, CRISPR and RNAi components, plasmids, etc. with protocols for their use. In addition, tools to facilitate neuroscience studies (e.g., species brain atlases; video protocols for performing...
neuronal recordings) and comparative work (e.g., tools for assessing gene homology, bioinformatics pipelines; sets of genetic markers) should be included along with the sequence databases currently available (marker datasets are available for most Mormyrid genera, but are not yet curated in a central repository).

3. Prospects for electric fish research – the next 30 years

Electric fish are a system with the potential to address many broad biological questions. Decades of research with traditional molecular tools on a variety of topics (Section 1) have been very productive; however, the development of molecular tools over the next decade or so (Section 2) will allow electric fish researchers to rapidly increase the pace of discovery. Applications of the genomic tools outlined above promise advances in electric fish research, and the ability to tackle some long-standing questions that have proven difficult to approach with pre-genomic tools. This section discusses a partial list of research avenues that we expect modern genomic tools to open up in the coming decades.

3.1. Gene expression profiles

First-generation sequencing approaches (Sanger sequencing of PCR products) to gene expression are usually limited to the study of one or a few genes simultaneously. Although valuable advances have been made with these techniques (Section 1), constructing a gene regulatory network using these methods is not practical. RNAseq can capture the entire expression profile of the target cells (Section 2.2), making it a formidable tool to describe gene networks and understand their consequences for tissue functions/development.

This strategy could determine whether compartmentalization of scn4aa and scn4ab expression occurs between subsets of muscle fibers in non-electrogenic fish (Zakon et al., 2008), and which gene interactions mediate such specialization. Likewise, single-cell RNA-seq (Kolodziejczyk et al., 2015) could identify candidate gene(s) responsible for signal transduction after comparing activated and non-activated electroreceptor cells. Comparative expression studies could also elucidate the molecular specializations in other electric fish tissues; for example, those between motoneurons and electromotoneurons.

Tissue comparisons have proven useful for understanding the differentiation between EO and skeletal muscle. Lamanna et al. (2014) found very similar gene expression profiles between muscle and EO transcriptomes from C. compressirostris. Further work has discovered hundreds of differentially expressed genes between these tissues, and found that the EO exhibits higher gene expression variability than muscle, across C. compressirostris, C. ishokwe, and Gnathonemus petersii (Lamanna et al., 2015). These comparisons have also identified potential genes involved in specialized EO functions, such as cell size, structure and insulation (Gallant et al., 2014), ion pumps, and membrane synthesis and turnover (Lamanna et al., 2015).

Complementing RNAseq of mRNA with miRNA-sequencing has broadened understanding of how gene expression profiles are regulated in the EO. An EO-specific miRNA (Traeger et al., 2015) and three muscle-expressed miRNA are upregulated in the EO compared to muscle (Pinch et al., 2016). To our knowledge, ribosome profiling (Ribo-seq) has yet to be used in electric fish research, but it could potentially provide information about post-transcriptional regulation. Ribo-seq captures the subset of mRNAs that are translated at a particular moment (Ingolia et al., 2009), and thus can be used to study the translational control of gene expression (Ingolia, 2014). Further application of these technologies should provide valuable insights to understand the function, evolution, and diversification of EOs and the nervous system control and perception of EODs.

3.2. Function of electrolyte proteins

The aforementioned work on the voltage-gated Na+ channel gene scn4aa (Section 1.2) describes an example of a protein adapted for electrolyte function. Yet, Na+ channels are not solely responsible for the propagation of action potentials, let alone the entire function of the electrolyte. There are likely additional examples of proteins co-opted for the electrolyte phenotype. Paralogs are prime candidates for such differential expression among tissues; nevertheless it would be most interesting to learn how cartilaginous electric fish (i.e., fishes that did not experience the teleost-specific whole genome duplication) became electrogenic.

Potential candidates of specialized electrolyte proteins could include those mentioned above (Section 3.1) as well as proteins involved in hormone/neurotransmitter reception (Sections 1.2 and 1.3 respectively). Perhaps the most industrious candidates should be K+ channels. Rashid and Dunn (1998) identified 19 potential K+ channel gene fragments in Apterodonus leptorhynchos, and suggested that K+ channels play critical roles in the precise regulation of neuronal firing that is required in complex nervous systems. Furthermore, rectifying K+ currents contribute to EOD production: K+ and Na+ kinetics are highly correlated in Sternoptygus, reflecting a high degree of coregulation between two distinct ion channels (McAnelly and Zakon, 2000), yet the expression K+ and Na+ channels is extremely tissue-specific in Eigenmannia virens (Ban et al., 2015). Therefore, it is worth exploring whether K+ channels have evolved in parallel to Na+ channels in Gymnotiform and Mormyrid electric fishes (Zakon et al., 2008). Whole-genome sequencing (see Section 2.2) could provide a comprehensive picture of the genes present and their homologies, suggest hypotheses about their evolutionary histories, and identify positively selected substitutions. The availability of well annotated genomes and transcriptomes would open the door to proteomic approaches; providing an opportunity to validate hypotheses in post-transcriptional regulation and alternative splicing. Additionally, predictions about proteins can be made from well annotated genomes/transcriptomes, allowing in silico trypsin-digested protein patterns to be generated as well as helping to identify proteins via mass spectrometry. The functional consequences of substitutions, alternative-splicing and post-transcriptional regulation could be inspected through the tools described in Sections 2.3 and 2.4.

3.3. Understanding the evolution of biodiversity using weakly electric fish as a model

With more NGS sequencing, electric fish may enable us to address some of the large-scale questions in evolutionary biology. Mormyrid and Gymnotiform weakly electric fish are remarkably speciose (Albert et al., 2005; Sullivan et al., 2000); their independent origins constitute a ‘natural experiment’ in species adaptation and radiation. In some electric fish clades, the changes involved in macroevolutionary diversification are mirrored at lower taxonomic levels; for example the repeated switch between bi- and tri-phasic EODs that seems to have occurred among genera and among conspecific populations in mormyrids (Arnegard and Carlson, 2005; Gallant et al., 2011). This is a truly rare opportunity to investigate potentially incipient speciation processes, with the aid of behavioral, life-history, and comparative genomics studies (Section 2.2).

Electric fish studies could address the relative importance – and potential conflict – between natural and sexual selection during the process of adaptive radiation (Andersson, 1994; Coyne and Orr, 2004; Schluter, 2001; Wagner et al., 2012). The EOD is used for species identification (Curtis and Stoddard, 2003; Feulner...
et al., 2009; Kramer, 1997; Silva et al., 2008) and therefore EOD divergence may be implicated in the radiation of these groups through sexual selection. Detailed morphological, physiological, and behavioral studies on electric fishes have given us an enviable understanding of their communication systems (Baker et al., 2013; Carlson, 2002; Carlson and Gallant, 2013). Two integrative studies suggest that adaptation to local ecological conditions drive divergence among Campylomormyrus species (Feulner et al., 2008); whereas selection on sexual signals seems more important among the Paramormyro (Arnegard et al., 2010). With genomic sequencing (Section 2.2.), it will be possible to use population genetic analyses to identify regions of the genome that indicate positive selection; i.e., genes/functions that are driving species divergence (Zakon et al., 2006). Comparisons of these regions between different species/populations of electric fish would be a valuable addition to the understanding of population genetics.

3.4. Understanding the role of cis-regulatory changes in the evolution of development

The availability of genomic-level data has opened up a number of broad-scale questions in evolution and development, in particular: the relative contributions of change in protein-coding sequence vs. cis-regulatory elements (Hoekstra and Coyne, 2007). Whole genome sequencing holds the key for unlocking the role cis-regulatory elements may play in the evolution, differentiation and maintenance of EOs (Section 2.2.). Zakon et al. (2008) suggest that the comparison of transcription factor binding sites from the gene scn4aa between electric and non-electrogenic taxa could help explain the genetic underpinnings of scn4aa specialization to electrocytes. We propose expanding on this with comparative genomics between fully sequenced genomes. RNAseq (Section 2.1) could tackle this question more broadly, by comparing additional regulatory elements, mechanisms, and genes involved in EO differentiation. We expect that a number of key signaling and regulatory molecules in electrocytes would have also evolved in concert with scn4aa and EOD signal divergence (Markham, 2013).

Once differences in cis-regulatory regions between genes/populations/species are identified, CRISPR (Section 2.4) can be utilized to assess their function. CRISPR can be directed to cis-regulatory regions, either to disrupt or replace the sequence (e.g., use HDR to replace an enhancer region for one gene with that of another gene, or of the same gene from a different species). CRISPR can be used to replace/alter transcription factor coding sequences, providing an opportunity to utilize electric fish as a model to address the relative impact of cis-regulation and coding sequence change in evolution (Hoekstra and Coyne, 2007; Schmidt et al., 2010).

3.5. Manipulation of neural circuits using transgenics

Numerous studies on electric fish have contributed to our current knowledge in systems and circuits neuroscience. Examples include: mechanisms for preservation and analysis of temporal information (Hopkins, 1999; Kawasaki, 1996), synaptic plasticity (Bell et al., 1997), cerebellar function (Sawtell and Bell, 2008), and effects of both motor (Bell and Grant, 1989) and neuromodulatory systems (Deemyad et al., 2013) on sensory processing. The neural circuitry underlying many complex electric fish behaviors has been well studied; the jamming avoidance response is completely understood from peripheral signal perception through motor output (Heiligenberg, 1991).

The electrosensory lobe is an important locus for understanding how neurons code information, i.e., how sensory perceptions are represented in configurations of all-or-none action potentials (Chacron et al., 2011; Clarke et al., 2015). Moreover, the repeated independent evolution of electrosensory systems in electric fishes allows for the rare opportunity of comparing convergent electrosensory adaptations to common pressures related to information processing or behavioral control (Kawasaki, 1996).

Further progress could be fueled by using advanced neuroscientific tools, such as the development of transgenic electric fish lines for optogenetics. This technique involves the expression of light-sensitive ion channel proteins (e.g., channelrhodopsin or halorhodopsin) in particular neuronal populations, and allows researchers to manipulate neuronal activity with light (Deisseroth, 2010; Mancuso et al., 2010). Optogenetics allows for monitoring and manipulation of neuronal activity from intact animals, in real time, with minimal interference of behavior. Optogenetics could also provide an alternate method (vs. voltage- or patch-clamp recordings) for recording ionic currents in EOs and/or the brain in species/situations where electrophysiology may be difficult.

Use of an inducible construct (e.g., tamoxifen-induced) to overexpress Caspase9 in specific neurons could facilitate studies of neuronal circuits. CRISPR could be used to direct, and insert, the inducible construct to the genome. Once integrated and induced, the overexpression of Caspase9 will lead to cell death via apoptosis (Druskovic et al., 2008), causing neuronal ablations. The comparison of pre- and post-treatment measurements could hint at the function of the lost brain region, as well as its role in behavior.

CRISPR can be used to generate ‘multi-transgenic’ fishes. For example, instead of tagging a single protein with GFP (see Section 2.4), other colored fluorescent proteins (e.g., RFP, YFP, BFP – see Dean and Palmer, 2014) can tag additional gene products, providing simultaneous comparisons of multiple proteins within a single fish (e.g., visualization of multiple neurotransmitter/hormones and their receptors in a single neuron, identification of neuronal subtypes/populations in whole brain sections). As mentioned in Section 2.4, in situ hybridization or fluorescent in situ hybridization in a ‘multi-transgenic’ fish can be used to label additional genes of interest that the transgenic fish doesn’t express.

Once genome manipulation is established in electric fish (Section 2.4), there are further transgenic manipulations that could be applied to answer precise questions. One such technique is ‘Brainbow’ labeling where (potentially) each cell in an organism displays a different fluorescent profile that is unchanged after cell division, during morphallaxis, and during/after development (Pan et al., 2013). CRISPR can be used to incorporate the ‘brainbow’ construct into the genome through HDR (Section 2.4). This technique has been used in skin to monitor regeneration (Chen et al., 2016), and in neurons to trace brain connections (Pan et al., 2011). The possibility to study neuronal connections in electric fish using this technique is very alluring. Many of these proposed questions/techniques are within the first accessible applications of this research.

Conclusion – looking to the future

The weakly electric fish community has a strongly integrated concept of phenotype (Fig. 1), which stems from extensive characterization of natural behavior (because of its historical roots in neuroethology). Both biologists and the groups of fishes that they study are diverse; the weakly electric fish community is rich with mechanistic understanding of phenotypes over multiple levels of biological analysis. Because of the new arrival of low cost, high throughput sequencing technologies, these rich sources of phenotypic data position the electric fish model as a unique system to tackle highly integrative 21st century biology research objectives, such as understanding the brain and integrating phenotype and genotype. A concerted effort by researchers in the electric fish community focused on strategic sequencing and techniques development over the next decade could reap outsized rewards, and
position the community as a leader in bringing genomics techniques to non-model systems.

Acknowledgements

The authors wish to thanks K. Dunlap, A. Charbonneau, D. Tack, A. Parigi and K. Allen for their helpful comments on an earlier version of this manuscript. The authors would additionally like to acknowledge the organizers of the 2016 Weakly Electric Fish Meeting in Montevideo, Uruguay, as well as its many participants, for ideas and feedback that formed the basis of this review. The authors are supported by NSF Grant 1455405 to JRG.


ghost knifefish (*Apteronotus leptorhynchus*): de novo assembly, annotation, and proteomics validation. BMC Genomics 16, 166.


