

GENOMIC BASIS OF ELECTRIC SIGNAL VARIATION IN AFRICAN WEAKLY ELECTRIC FISH

By

Mauricio Losilla-Lacayo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Zoology—Doctor of Philosophy
Ecology, Evolutionary Biology and Behavior—Dual Major

2021

ABSTRACT

GENOMIC BASIS OF ELECTRIC SIGNAL VARIATION IN AFRICAN WEAKLY ELECTRIC FISH

By

Mauricio Losilla-Lacayo

A repeated theme in speciation is reproductive isolation centered around divergence in few, highly variable traits, specially in cases without strong geographic isolation and high speciation rates. Understanding the genomic basis of highly variable traits that are key to speciation is a major goal of evolutionary biology, because they can characterize crucial drivers and foundations of the speciation process. African weakly electric fish (Mormyridae) are a decidedly speciose clade of teleost fish, and their electric organ discharges (EODs) are highly variable traits central to species divergence. However, little is known about the genes and cellular processes that underscore EOD variation. In this dissertation, I employ RNAseq and Nanopore sequencing to study the genomic basis of electric signal variation in mormyrids. In Chapter 1, I take a transcriptome-wide approach to describe the molecular basis of electric signal diversity in species of the mormyrid genus *Paramormyrops*, divergent for EOD complexity, duration and polarity. My results emphasize genes that influence the shape and structure of the electrocyte cytoskeleton, membrane, and extracellular matrix, and the membrane's physiological properties. In Chapter 2, I compare gene expression patterns between electric organs that produce long vs short EODs. The results strongly support known aspects of morphological and physiological bases of EOD duration, and for the first time I identified specific genes and broad cellular processes expected to that alter morphological and physiological properties of electrocytes, most striking among these is the differential expression of multiple potassium voltage-gated channels. These two chapters independently identified the gene *epdl2* as of interest for EOD divergence. In Chapter 3, I study the molecular evolutionary history of *epdl2* in Mormyridae, with emphasis on *Paramormyrops*. My results suggest that three rounds of gene duplication produced four *epdl2* paralogs in a *Paramormyrops* ancestor. In

addition, I identify ten sites in *epdl2* expected to have experienced strong positive selection in paralogs and implicate them in key functional domains. Overall, the results of this dissertation greatly solidify and expand our understanding of how the genome underpins changes to electrocytes, and in turn, divergence in their electric signals, a highly variable trait that may facilitate speciation in African weakly electric fish. This work provides an evidence-grounded list of candidate genes for functional analyses aimed to corroborate their contribution to the EOD phenotype.

ACKNOWLEDGEMENTS

PhD dissertations in Biology have become massive undertakings that take many years to complete. They tacitly require collaborative efforts that go beyond the student and guidance committee, plus constant help and encouragement from support networks. My dissertation has been no exception, I owe its successful completion to a great many people.

First, I will like to thank my guidance committee, especially Dr. Jason Gallant, my advisor. Jason accepted me into his lab when I found myself without a lab after my first year. He provided financial support through part of my Summer stipends and all expenses related to data collection. More importantly, his input was fundamental in every experiment and analysis I performed. Likewise, my committee members Dr. Ingo Braasch, Dr. Kay Holekamp, and Dr. Kim Scribner have been instrumental for the success of my research projects. They were always available when I needed their input and were keen to provide constructive feedback time after time throughout my work. I deeply appreciate your input in shaping this dissertation, and my professional development in general.

One of the perks of my time spent in the Gallant Lab is the terrific group of labmates I was lucky to have: Dr. Savvas Constantinou, Hope Healey, Lauren Koenig, Dr. David Luecke, Dr. Sophie Picq, and Jared Thompson. They sit at the intersection of people who provided scientific assistance and invaluable friendship. I was also fortunate to attend the lab meetings and other activities from the Braasch lab and the Ganz lab (Dr. Julia Ganz). They also provided important suggestions to my work, especially Dr. Andrew Thompson, who always had time and disposition to talk about my projects. I was a teaching assistant for several semesters, and I had the pleasure of working with Dr. Teresa McElhinny. She is fantastic, and taught be invaluable lessons about teaching. My third chapter would not have been possible without sequences kindly facilitated by Rose Peterson and Dr. Stacy Pirro, samples generously donated by Dr. Bruce Carlson and Erika Schumacher, and specimens captured in a lab field trip to

Gabon, 2019. I was not part of this field trip, but it enjoyed the precious collaboration of Hans Mipounga, Franck Nzgiou, and Nestor Ngoua.

I could fill an entire page with the names of wonderful friends I have made during my time in the Greater Lansing Area. I was lucky to become a part of communities and support networks composed of friends from all over the world. When I started my PhD, I never imagined I was going to meet so many outstanding people. Thank you all so much for the great times we have shared, and I hope there are more moments like that on our futures. Sadly, some of these friends passed away during these years. I am shattered by your absence, yet I am grateful for the time shared with each of you.

My deepest, most heartfelt acknowledgment is to my family. This dissertation would not have been possible without your unwaning support. My beloved wife, Nati, has done so much for me, at so many levels, that I can't find the words to describe my gratitude and appreciation. Every day is a good day because we are together. Thank you.

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
GENERAL INTRODUCTION.....	1
Background: Understanding the Genetic Basis of Phenotypic Variation.....	1
An Introduction to the Mormyrid Study System	3
Dissertation Overview.....	4
REFERENCES.....	7
CHAPTER 1: THE TRANSCRIPTIONAL CORRELATES OF DIVERGENT ELECTRIC ORGAN DISCHARGES IN <i>PARAMORMYROPS</i> ELECTRIC FISH.....	13
Introduction	13
Methods.....	16
Sample collection	16
RNA extraction, cDNA library preparation and Illumina Sequencing	16
Read processing and data exploration.....	17
Data Analysis.....	17
Results.....	20
Overall Results	20
Set A: Differential Expression Analysis.....	20
Set B: Expression Based Clustering	21
Set C: Intersection of Phylogenetically Informative Comparisons and Expression Based Clustering	21
Discussion.....	22
Waveform Duration	23
Waveform Polarity	26
Waveform Complexity	27
Conclusions	29
APPENDIX.....	32
REFERENCES.....	49
CHAPTER 2: MECHANISTIC INSIGHTS INTO GENE EXPRESSION CHANGES AND ELECTRIC ORGAN DISCHARGE ELONGATION IN MORMYRID ELECTRIC FISH	57
Introduction	57
Methods.....	60
Study fish.....	60
Experimental conditions	60
EOD recordings	61
Experimental treatments.....	61
Tissue dissections, RNA extraction and sequencing	62
Read processing and data exploration.....	63
Differential Gene Expression	63
Functional enrichment analysis	64

Differentially expressed genes of highest interest for EOD duration	65
Results	65
EOD duration responses to hormone treatment	65
RNAseq data and differentially expressed genes	66
Functional enrichment analysis	66
Differentially expressed genes of highest interest for EOD duration	68
Discussion	68
EOD Duration and gene expression in the electric organ	69
Functional enrichment analysis	70
Early Changes in Response to 17 α MT	70
Late Changes in Response to 17 α MT	71
Differentially expressed genes of highest interest for EOD duration	72
Genes that affect electrocyte morphology	72
Genes that affect electrocyte physiology	73
Conclusions	76
APPENDIX	78
REFERENCES	90
 CHAPTER 3: MOLECULAR EVOLUTION OF THE EPENDYMIN-RELATED GENE <i>EPDL2</i> IN AFRICAN WEAKLY ELECTRIC FISH	96
Introduction	96
Methods	99
<i>epdl2</i> sequences from Osteoglossiformes genomes	99
Confirm <i>epdl2</i> triplication in <i>Paramormyrops kingsleyae</i>	99
<i>epdl2</i> paralogs in the <i>P. kingsleyae</i> Nanopore assembly	99
PCR amplification and Sanger sequencing of <i>epdl2</i> paralogs in <i>P. kingsleyae</i>	99
<i>epdl2</i> sequences across Mormyridae	100
Primer design	100
DNA extraction, PCR amplification and <i>epdl2</i> sequencing	100
Analysis of Nanopore reads	101
Read processing	101
Identification of <i>epdl2</i> genes	102
<i>epdl2</i> gene tree and <i>epdl2</i> duplication history	103
Selection Tests	103
Structural predictions of <i>EpdI2</i>	105
Results	105
<i>Paramormyrops kingsleyae</i> has three complete <i>epdl2</i> paralogs	105
<i>epdl2</i> sequences and copy number across Mormyridae	106
<i>Paramormyrops</i> underwent three <i>epdl2</i> duplications	106
Selection Tests	108
Structural predictions of <i>EpdI2</i>	110
Discussion	111
<i>epdl2</i> amplicon sequencing	111
<i>Paramormyrops</i> underwent three <i>epdl2</i> duplications	112
Selection Tests	113
Functional Consequences of Molecular Evolution in <i>epdl2</i> Paralogs	115
Conclusions	118
APPENDIX	120

REFERENCES..... 139

LIST OF TABLES

Table 1.1. Phenotypic and collection information of the samples studied.	33
Table 1.2. All ten possible pairwise DGE comparisons with the total number of DEG and enriched GO terms for each. Also indicated is whether each comparison is informative for contrasting each EOD feature. The phenotypes for waveform polarity can only be contrasted in comparisons where both OTUs have penetrations. Informative comparisons for each EOD feature (Set A') are marked with an * in the column of the EOD feature they contrasted.....	34
Table 1.3. Total number of upregulated genes and enriched GO terms in Set C for each EOD feature, phenotype and ontology.....	35
Table 1.4. Selected DEG in Set C for waveform duration by “general” functional class and EOD phenotype, and highlights of their predicted function.....	36
Table 1.5. Selected DEG in Set C for waveform polarity, by “general” functional class and EOD phenotype, and highlights of their expected function.	38
Table 1.6. Selected DEG in Set C for waveform complexity, by “general” functional class and EOD phenotype, and highlights of their expected function.	41
Table 2.1. All three possible pairwise differential gene expression comparisons with the number of genes expressed, differentially expressed, and their breakdown by treatment	79
Table 2.2. Genes differentially expressed in the three pairwise treatment contrasts that also belong to the key gene sets	80
Table 2.3. Select differentially expressed genes from the three pairwise treatment contrasts. Shown are the number of genes in each general theme and featured gene examples from each general theme.....	81
Table 3.1. Source of epdI2 sequences, number and size of epdI2 genes found in every species studied	121
Table 3.2. Primers used to amplify and sequence epdI2 paralogs across Mormyridae	123
Table 3.3. PCR reagents and final concentrations used to amplify each epdI2 paralog in <i>P. kingsleyae</i> .	124
Table 3.4. PCR conditions used to amplify each epdI2 paralog in <i>P. kingsleyae</i>	124
Table 3.5. Mormyrid species and their NCBI bioproject codes that guided epdI2 primer design.....	125
Table 3.6. PCR reagents and final concentrations used to amplify all epdI2 paralogs across Mormyridae	126
Table 3.7. PCR conditions used to amplify all epdI2 paralogs across Mormyridae	126

Table 3.8. Sites along *epdl2* that have experienced positive selection in the osteoglossiform *epdl2* gene tree and have evolved at increased ω rates in duplicated vs non-duplicated mormyrid lineages. Branches supported by each EBF value are marked by rectangles in Figure 3.10 (EBF values for a given site are arranged to match their corresponding branches from top to bottom). 127

Table 3.9. Summary of the amino acid substitutions observed in the *Epdl2* paralogs at the ten sites under positive selection and increased ω rates in duplicated vs non-duplicated mormyrid lineages 127

LIST OF FIGURES

Figure 1.1. Electric organ discharge (EOD) diversity and electric organ anatomy in *Paramormyrops*. EOD traces from specimens in this study and representative parasagittal sections of the five *Paramormyrops* operational taxonomic units (OTUs) considered in this study. 200x magnification on *P. kingsleyae* EODs reveals a P0 phase on triphasic EODs only. Individuals with triphasic EODs all have penetrations, whereas individuals with biphasic EODs do not. OTUs with ‘inverted’ polarity triphasic EODs have large penetrations compared to OTUs with normal polarity triphasic EODs. Ant. = anterior, C = connective tissue septa, N = nerve, NPp = non-penetrating, posteriorly innervated stalks, M = microstalklets (profusely branched stalks), P = penetrations, Pa = penetrating, anteriorly innervated stalks, Post. = posterior, S = stalks. Image from *P. kingsleyae* biphasic originally appeared in Gallant et al. 2011. 43

Figure 1.2. Diagram of how we constructed the lists of upregulated genes of Set C. DEG = differentially expressed genes, N = number of comparisons made for each set, OTU = operational taxonomic unit 44

Figure 1.3. Heatmap of sample by sample correlations in gene expression, and the inferred relationships among OTUs from these expression correlation values. OTU = operational taxonomic unit. 45

Figure 1.4. Set C for waveform duration. A) Expression patterns of Set C genes for the waveform duration phenotypes short EODs (purple background) and long EODs (yellow background). Samples are sorted alphabetically on the X axis. The lines connect transformed gene expression values across all samples; light-color lines represent one gene, the dark-color line is the average expression pattern of all genes. B) Gene Ontology (GO) terms for Biological Process and Cellular Component found enriched in the gene lists from (A). The X axis shows transformed p-values, the longer a bar the smaller its p-value. The direction and color of a bar indicate the phenotype in which the GO term is enriched [same color code as (A)]. 46

Figure 1.5. Set C for waveform polarity. A) Expression patterns of Set C genes for the waveform polarity phenotypes small penetrations (red background) and large penetrations (grey background). Samples are sorted alphabetically on the X axis. The lines connect transformed gene expression values across all samples; light-color lines represent one gene, the dark-color line is the average expression pattern of all genes. B) Gene Ontology (GO) terms for Biological Process and Cellular Component found enriched in the gene lists from (A). The X axis shows transformed p-values, the longer a bar the smaller its p-value. The direction and color of a bar indicate the phenotype in which the GO term is enriched [same color code as (A)]. Pen = penetrations. 47

Figure 1.6. Set C for waveform complexity. A) Expression patterns of Set C genes for the waveform complexity phenotypes triphasic (orange background) and biphasic (blue background). Samples are sorted alphabetically on the X axis. The lines connect transformed gene expression values across all samples; light-color lines represent one gene, the dark-color line is the average expression pattern of all genes. B) Gene Ontology (GO) terms for Biological Process and Cellular Component found enriched in the gene lists from (A). The X axis shows transformed p-values, the longer a bar the smaller its p-value. The direction and color of a bar indicate the phenotype in which the GO term is enriched [same color code as (A)]. 48

Figure 2.1. Generalized model of a biphasic EOD and its underlying APs. Top: APs of the posterior (AP1, red) and anterior (AP2, blue) faces of the electrocyte, and the resulting EOD (black, calculated as AP1 – AP2). Bottom: Cartoon electrocytes (posterior face red, anterior face blue) illustrating the Na ⁺ and K ⁺ ionic currents at key moments in the EOD. Adapted from (Stoddard and Markham 2008).	82
Figure 2.2. Photographs of the experimental fish tanks' configuration. A top view (left) shows the housing arrangement attached to the middle of the tank's bottom, a frontal view (right) illustrates electrode positioning and attachment during recording sessions.	83
Figure 2.3. EOD duration at the onset (left) and at the end (right) of the experiment. The circle and vertical lines represent the mean EOD duration +/- standard deviation per treatment, open triangles are per fish measurements. A small horizontal jitter was added to better visualize overlapping observations. Each asterisk represents significant differences (p < 0.001) between the treatments connected by its overlying horizontal bar.	84
Figure 2.4. EOD duration per treatment throughout the experiment. Each colored circle and its vertical lines represent the mean EOD duration +/- standard deviation per treatment and day. A small horizontal jitter was added to better visualize overlapping values. Days -4 to 0 are part of the acclimation period, treatment-specific manipulations were performed on Day 0 after taking EOD recordings.	85
Figure 2.5. Change in EOD duration in one representative fish per treatment. Each graph shows the traces of the averaged EODs from individual fish recordings taken on Days 0, 1, and 8 (not applicable for treatment T1day). When EOD duration changes are very small, traces from more recent days overlap older traces.	86
Figure 2.6. Heatmap of sample by sample correlations in gene expression, and the inferred relationships among treatments from these expression correlation values.	87
Figure 2.7. Cartoon depiction of how gene expression changes detected in the three pairwise treatment comparisons relate to each other in terms of time of exposure to 17 α MT.	87
Figure 2.8. Heatmap of mitch enrichment scores (s) for all (n = 39) significantly enriched gene sets (higher dimensional enrichment score > 0.1, FDR p < 0.01) in the three pairwise treatment contrasts studied. Gene sets are further classified into general categories (legend).	88
Figure 2.9. Simulated effects on EOD amplitude and duration of delayed depolarizations of the anterior electrocyte face in relation to the posterior face. The AP of each face is simulated as a normal distribution of sd = 1. Threshold for EOD detection = 0.004. A) EOD amplitude (top) and duration (bottom) across delays ranging from 0 to 1. Units and values are irrelevant. B) a representative EOD plot with delay = 0.50. Green = AP of posterior face, red = AP of anterior face, blue = resulting EOD, black horizontal lines = EOD detection threshold.	89
Figure 3.1. Graphical summary of the bioinformatic pipeline we leveraged to identify epd12 genes in the filtered amplicons from each species, see Methods for details. Same-colored connector arrows represent the analysis with a specific value for c (cd-hit's sequence identity threshold parameter, range analyzed 0.84-0.91). Magenta connector arrows represent the analysis with the c value chosen with the selection criteria.	128

Figure 3.2. Genomic region with the <i>epdl2</i> paralogs in <i>Paramormyrops kingsleyae</i> . DNA sequence is represented by the black line, numbers above indicate base positions, annotations are shown below the line. Flanking genes are depicted in cyan, <i>epdl2</i> paralogs in red (start to stop codon) and yellow (CDSs). Arrows direction indicate gene membership to a strand (right for +, left for -). Magenta blocks mark the Sanger-sequenced regions.....	128
Figure 3.3. Cleaned PCR products for each <i>Paramormyrops kingsleyae</i> <i>epdl2</i> paralog. Lanes with a PCR product are labeled with the respective abbreviated paralog name. The unlabeled lane (lane 3) holds a DNA ladder, its size legend is shown inserted	129
Figure 3.4. Mormyrid <i>epdl2</i> gene tree. Bootstrap support values > 60% are shown. <i>epdl2</i> paralogs are classified based on our best hypothesis of <i>epdl2</i> duplications	130
Figure 3.5. Broad phylogenetic relationships predicted among the studied species from the <i>Paramormyrops</i> + <i>Marcusenius ntemensis</i> clade. Based on the AFLP phylogeny from (Sullivan et al. 2004)	131
Figure 3.6. Topology of the <i>epdl2</i> gene tree with duplicated and non-duplicated (single copy) mormyrid lineages indicated. This branch partition was used in the RELAX and Contrast-FEL tests.	132
Figure 3.7. Topology of the <i>epdl2</i> gene tree with branches tested for positive selection with aBSREL indicated. Branches with significant results are labeled A and B.	133
Figure 3.8. Sites along <i>epdl2</i> that have experienced positive selection (MEME, $p < 0.1$) in the osteoglossiform taxa studied. p values have been transformed so that higher values on the y axis represent lower p values. Horizontal line marks the significance threshold and significant sites have been labeled.....	134
Figure 3.9. Sites along <i>epdl2</i> with higher ω values in duplicated vs non-duplicated mormyrid lineages (Contrast-FEL, $p < 0.1$). p values have been transformed so that higher values on the y axis represent lower p values. Horizontal line marks the significance threshold and significant sites have been labeled.	134
Figure 3.10. Topology of the <i>epdl2</i> gene tree, we highlight branches and sites where we detected signals of selection. Orange branches are recently duplicated <i>epdl2</i> paralogs, selection has intensified in these branches relative to the non-duplicated mormyrid lineages. Branches with asterisks experienced positive selection. Colored rectangles are labeled with sites along <i>epdl2</i> where selection has intensified in duplicated lineages and have experienced positive selection. These rectangles are placed on the branches where exploratory evidence suggests they underwent positive selection. All rectangles (purple) but one (green) map to the duplicated lineages.	135
Figure 3.11. <i>Paramormyrops szaboi</i> <i>Epdl2</i> predicted amino acid sequence, annotated with salient structural properties. Pink: signal peptide; orange: conserved cysteine residues forming disulfide bonds; purple: N-glycosylation sites; blue: positively selected sites with increased ω rates in <i>epdl2</i> paralogs, number labels indicate sites' position in the osteoglossiform alignment of <i>epdl2</i> CDSs. Yellow arrows and red blocks are β strands and α helices interpreted by Geneious from the 3D model.....	136
Figure 3.12. Predicted 3D structure of <i>Paramormyrops szaboi</i> <i>Epdl2</i> . A-D: four different views superimposed on its best structural analog, <i>Xenopus tropicalis</i> <i>epdr1</i> (purple backbone trace). E-F: two	

different views highlighting the location of structural features (disulfide bonds yellow, N-glycosylation sites green) and residues of interest (colored sticks): conserved proline residue pink, and our ten sites of highest interest (105-106 red, 125-7, 129 blue, 150, 172, 176 cyan, 153 orange). Pink, red and orange residues are likely oriented towards the hydrophobic pocket (E), blue and cyan residues are probably directed towards the exterior (F)..... 137

Figure 3.13. Theoretical gene trees with the topological patterns expected under two duplication scenarios. Arrows indicate gene duplications, same-colored arrows mark gene duplications stemming from the same duplication event. A: Three duplications scenario, each duplication yields one additional paralog. B: two duplications scenario, each duplication doubles the number of paralogs..... 138

GENERAL INTRODUCTION

Background: Understanding the Genetic Basis of Phenotypic Variation

Variation in phenotypic traits is a key component in the speciation processes. This is perhaps most evident in taxa undergoing adaptive radiations and explosive diversifications (sensu Givnish 2015), where variation in a few hypervariable traits is responsible for rapid speciation. Darwin's finches radiated based on their beak morphology and feeding habits (Grant and Grant 2003), African cichlid fish vary mainly in their external jaw morphology and male nuptial color (Kocher 2004), and Hawaiian swordtail crickets (*Laupala*) diversified around their male calling song (Shaw 2000) and cuticular lipids (Mullen et al. 2007). These hypervariable traits have a variety of key functions, from opening ecological niches (e.g. finches' beaks and cichlids' jaws) to attracting and selecting mates (e.g. crickets' calling song, cichlids' nuptial color, and their females' preferences). However, phenotypic characters often influence fitness in multiple, complex and coordinated means, especially when they are constituents of behaviors; for example, the beak morphology of Darwin's finches affects their male song (Podos 2001; Grant and Grant 2008). This complicates parsing the evolutionary forces that shape phenotypic divergence. Moreover, it is not unusual for some traits, so called 'magic traits', to be targets of divergent selection and to pleiotropically contribute to non-random mating (Servedio et al. 2011). In fact, natural and sexual selection interact frequently during diversification processes (Servedio and Boughman 2017).

Although not as well characterized as the overlying phenotypic variation, the study of the genomic causes of these 'hypervariable' traits has advanced substantially. For some traits these causes are relatively well identified, for example, the genomic regions around the *alx1* and *hmga2* loci, which are associated with variation in beak shape and size, respectively, are strongly indicated in Darwin's finches divergence (Lamichhaney et al. 2015, 2016; Han et al. 2017). More often, the genetic basis of these traits is more complicated, particularly when mate selection is clearly involved. For example: in Hawaiian crickets, male calling song and female preference are both polygenic traits, and the genetic loci that control their variation show linkage disequilibrium (Shaw and Lesnick 2009; Wiley et al. 2012).

Contrary to expectations, at least one of these QTLs is not in a region of low recombination (Blankers et al. 2018), thus the evidence is still consistent with potentially pleiotropic loci influencing male song and female preference. In the example of African cichlid fish, male nuptial color may be controlled by a small (Magalhaes and Seehausen 2010; O’Quin et al. 2012) or large (Ding et al. 2014) number of epistatic loci, whereas female mating preferences are influenced by few loci (Haesler and Seehausen 2005), which are likely unlinked from those that control male nuptial color (Ding et al. 2014). A mapping study reported 41 QTLs and 13 epistatic interactions that influence chromatophore coloration in these fish and identified some genes from the *pax* gene family as candidate genes for this trait (Albertson et al. 2014). Furthermore, African cichlid fish likely benefited from exceptional features of their genomic architecture that facilitated their unique diversification. They have an excess of gene duplications and standing genetic variation, underwent accelerated evolution of regulatory and coding sequences, and possess novel microRNAs and transposable element insertions that likely influence divergent gene expression (Brawand et al. 2014).

Despite the complexity illustrated by these examples, understanding the genomic bases of phenotypic diversity is a major goal of evolutionary biology. Linking genotype to phenotype in this manner facilitates the study of numerous aspects of evolutionary model systems, including the mutational effects on phenotypes, the genetic changes that underlie phenotypic convergence, the roles of pleiotropy, epistasis and genetic linkage on phenotypic evolution, the importance of changes in coding or regulatory regions, and the contributions of standing genetic variation and new mutations (Peichel and Marques 2017). Adaptive radiations and explosive diversifications offer exceptional advantages to study the genomic bases of phenotypic diversity. They can provide replication under a controlled phylogenetic framework (Seehausen 2006), and they couple ample phenotypic differentiation with relatively “cleaner” genomic signals between recently diverged species (Hulseay 2009). Younger or ongoing speciation events offer the best opportunities, as they constitute the initial stages of

diversification, where interspecific differences are still small and phenotypic variation may be present at population levels.

An Introduction to the Mormyrid Study System

Although Osteoglossiformes are an ancient, modestly sized clade of teleost fish, it contains one of the most rapidly speciating groups of ray-finned fishes, the African weakly electric fish (Teleostei: Mormyridae) (Carlson et al. 2011; Rabosky et al. 2013). Mormyrid diversification is partly driven by the adaptive radiations or explosive diversifications in *Campylomormyrus* (Feulner et al. 2007, 2008) and *Paramormyrops* (Sullivan et al. 2002, 2004). The latter has produced more than 20 species in the basins of West-Central Africa (Lavoué et al. 2008), mostly within the last 500 000 years (Sullivan et al. 2002).

Mormyrids success rests upon their capacity for electrolocation and electroreception: they can generate weakly electric discharges upon the surrounding water, and they can detect and analyze the resulting electric fields (Alves-Gomes 2001). These electric discharges, termed electric organ discharges (EODs), are a behavior with a dual role in electrolocation (Lissmann and Machin 1958; von der Emde et al. 2008) and intraspecific communication (Möhres 1957; Kramer 1974). They have facilitated the exploitation of nocturnal environments and has bestowed mormyrids with an almost private communication channel. EODs are also a discrete electrophysiological event with a well-understood morphological and neurophysiological basis (Carlson 2002; Caputi et al. 2005). They are generated by specialized cells (electrocytes) that constitute the electric organ (EO), located in the caudal peduncle (Hopkins 1986). Mormyrid EOs are comprised of 80-360 electrocytes (Bass 1986), and an individual EOD is produced by their synchronous discharge (Bennett and Grundfest 1961).

Mormyrid EOD waveforms vary primarily in their complexity, polarity, and duration (Hopkins 1999b; Sullivan et al. 2002). Waveform complexity refers to the number of phases (peaks) present in the EOD, and it is a product of the ultrastructural configuration of the electrocytes (for details see Bennett and Grundfest 1961; Szabo 1961; Bennett 1971; Hopkins 1999b; Gallant et al. 2011; Carlson and Gallant

2013). Waveform polarity describes alterations to the waveform shape caused when there is a large, initial, negative phase (Bennett and Grundfest 1961; Arnegard et al. 2005; Gallant et al. 2011). EOD duration, the time it takes the EOD to complete, is the feature that varies the most between species (Arnegard and Hopkins 2003; Gallant et al. 2011), up to 100x (Hopkins 1999a), and is the basis for recognition of species-specific signals (Hopkins and Bass 1981; Arnegard et al. 2006).

EODs, exhibit little intraspecific variation, are effective mechanisms of prezygotic isolation (Hopkins and Bass 1981; Arnegard et al. 2006), and often are the most reliable method for identifying species (Hopkins 1981; Bass 1986; Alves-Gomes and Hopkins 1997; Sullivan et al. 2000; Arnegard et al. 2005).

This is particularly evident in *Paramormyrops* (Picq et al. 2020), where EOD waveform evolution greatly outpaces that of morphology, size, and trophic ecology (Arnegard et al. 2010). All this suggests that EOD divergence may be central to the speciation process in these fishes. Therefore, mormyrids in general, and *Paramormyrops* in particular, represent a unique opportunity to study the genomic basis of a hypervariable phenotypic trait (the EOD) across multiple stages of diversification and repeated speciation events.

Dissertation Overview

Previous work has provided valuable information about the similarities and differences in gene expression between EOs and skeletal muscles in independently evolved electric fish (Gallant et al. 2014; Traeger et al. 2015). However, little is known about the genes and cellular processes that underscore EOD divergence. Genes related to metabolism of fatty acids, ion transport, and neuronal function (Lamanna et al. 2015); and several ion channel genes (Nagel et al. 2017); have been observed differentially expressed between *Campylomormyrus* species. Here, I employed RNAseq and Nanopore sequencing to study the genomic basis of electric signal variation in mormyrids and *Paramormyrops*.

In Chapter 1, I took a transcriptome-wide approach to describe the molecular basis of electric signal diversity in *Paramormyrops* species divergent for EOD complexity, duration and polarity. I used

RNA-sequencing to comprehensively examine differential gene expression in the adult EOs of five *Paramormyrops* operational taxonomic units (OTUs), and thus identify gene expression correlates of each of the three main EOD waveform features of electric signal diversity. My results emphasize genes that influence the shape and structure of the electrocyte cytoskeleton, membrane, and extracellular matrix (ECM), and the membrane's physiological properties, to exhibit predictable differences between *Paramormyrops* OTUs with divergent EOD phenotypes. A particularly interesting result was that of the gene *epdl2*: I found three copies in the *Paramormyrops kingsleyae* reference genome, whereas only one copy is present in related species with genomic resources available; and gene expression of two of these paralogs correlated with variation on EOD waveform complexity.

In Chapter 2, I leverage on the fact that EOD duration can be modulated by androgen hormones (Bass and Hopkins 1983, 1985; Bass et al. 1986; Bass and Volman 1987; Freedman et al. 1989; Herfeld and Moller 1998) to compare gene expression patterns between EOs producing long vs short EODs. I experimentally applied 17 α -methyltestosterone to individuals from the mormyrid species *Brienomyrus brachyistius* over an 9-day experimental period, and used RNAseq to examine patterns of differential gene expression over the course of this treatment compared to control organisms. This experiment was truly unique because it afforded comparisons between conspecific individuals, hence controlling for otherwise inevitable species-depending confounding factors. The results strongly support known aspects of morphological and physiological bases of EOD duration, and for the first time I identified specific genes and broad cellular processes that alter morphological and physiological properties of electrocytes during seasonally plastic EOD changes, most striking among these is the differential expression of multiple K⁺ voltage-gated channels.

Finally, in Chapter 3, I followed on the gene *epdl2*, flagged in Chapter 1 as of interest for EOD divergence. I elucidated the molecular evolutionary history of *epdl2* in Mormyridae, with emphasis on *Paramormyrops*. First, I confirmed the existence of three *epdl2* paralogs in *P. kingsleyae*. Then I

employed multiplexed amplicon Nanopore sequencing to identify *epdl2* gene copies across *Paramormyrops* and Mormyridae. I elucidated paralog diversity, inferred duplications history, and tested for signatures of natural selection on branches and sites. Lastly, I modeled the 3D structure of a representative *epdl2* protein and proposed how the sites deemed under the strongest positive selection in paralogs may affect protein function.

Overall, the results of this dissertation greatly solidify and expand our understanding of how the genome underpins changes to electrocytes, and in turn, divergence in their electric signals, a hypervariable trait that may facilitate speciation in African weakly electric fish. This work provides an evidence-grounded list of candidate genes for functional analyses aimed to corroborate their contribution to the EOD phenotype.

REFERENCES

REFERENCES

- Albertson, R. C., K. E. Powder, Y. Hu, K. P. Coyle, R. B. Roberts, and K. J. Parsons. 2014. Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. *Mol. Ecol.* 23:5135–5150.
- Alves-Gomes, J. A. 2001. The evolution of electroreception and bioelectrogenesis in teleost fish: a phylogenetic perspective. *J. Fish Biol.* 58:1489–1511.
- Alves-Gomes, J., and C. D. Hopkins. 1997. Molecular insights into the phylogeny of mormyrid fishes and the evolution of their electric organs. *Brain Behav. Evol.* 49:324–350.
- Arnegard, M. E., S. M. Bogdanowicz, and C. D. Hopkins. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. *Evolution* 59:324–343.
- Arnegard, M. E., and C. D. Hopkins. 2003. Electric signal variation among seven blunt-snouted *Brienomyrus* species (Teleostei: Mormyridae) from a riverine species flock in Gabon, Central Africa. *Environ. Biol. Fishes* 67:321–339.
- Arnegard, M. E., B. S. Jackson, and C. D. Hopkins. 2006. Time-domain signal divergence and discrimination without receptor modification in sympatric morphs of electric fishes. *J. Exp. Biol.* 209:2182–2198.
- Arnegard, M. E., P. B. McIntyre, L. J. Harmon, M. L. Zelditch, W. G. R. Crampton, J. K. Davis, J. P. Sullivan, S. Lavoué, and C. D. Hopkins. 2010. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *Am. Nat.* 176:335–356.
- Bass, A. H. 1986. Species differences in electric organs of mormyrids: Substrates for species-typical electric organ discharge waveforms. *J. Comp. Neurol.* 244:313–330.
- Bass, A. H., J.-P. Denizot, and M. A. Marchaterre. 1986. Ultrastructural features and hormone-dependent sex differences of mormyrid electric organs. *J. Comp. Neurol.* 254:511–528.
- Bass, A. H., and C. D. Hopkins. 1985. Hormonal control of sex differences in the electric organ discharge (EOD) of mormyrid fishes. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 156:587–604.
- Bass, A. H., and C. D. Hopkins. 1983. Hormonal control of sexual differentiation: changes in electric organ discharge waveform. *Science* 220:971–974.
- Bass, A. H., and S. F. Volman. 1987. From behavior to membranes: testosterone-induced changes in action potential duration in electric organs. *Proc. Natl. Acad. Sci. U.S.A.* 84:9295–9298.
- Bennett, M. V. L. 1971. Electric Organs. Pp. 347–491 in W. S. Hoar and D. J. Randall, eds. *Fish Physiology*. Academic Press, London.
- Bennett, M. V. L., and H. Grundfest. 1961. Studies on the morphology and electrophysiology of electric organs. III. Electrophysiology of electric organs in mormyrids. Pp. 113–35 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, Amsterdam.

Blankers, T., K. P. Oh, A. Bombarely, and K. L. Shaw. 2018. The Genomic Architecture of a Rapid Island Radiation: Recombination Rate Variation, Chromosome Structure, and Genome Assembly of the Hawaiian Cricket *Laupala*. *Genetics* 209:1329–1344. *Genetics*.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, J. Turner-Maier, J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J. F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. A. Conte, H. D’Cotta, O. Eshel, L. Gaffney, F. Galibert, H. F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. A. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, A. P. Lee, I. MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga, S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Yin, N. Okada, T. D. Kocher, E. A. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen, and F. Di Palma. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513:375–381.

Caputi, A. A., B. A. Carlson, and O. Macadar. 2005. Electric Organs and Their Control. Pp. 410–451 *in* T. H. Bullock, Hopkins C.D., Popper A.N., and Fay R.R., eds. *Electroreception*. Springer Handbook of Auditory Research, vol 21, New York.

Carlson, B. A. 2002. Electric signaling behavior and the mechanisms of electric organ discharge production in mormyrid fish. *J. Physiol. Paris* 96:405–419. Elsevier.

Carlson, B. A., and J. R. Gallant. 2013. From sequence to spike to spark: evo-devo-neuroethology of electric communication in mormyrid fishes. *J. Neurogenet.* 27:106–129.

Carlson, B. A., S. M. Hasan, M. Hollmann, D. B. Miller, L. J. Harmon, and M. E. Arnegard. 2011. Brain evolution triggers increased diversification of electric fishes. *Science* 332:583–586.

Ding, B., D. W. Daugherty, M. Husemann, M. Chen, A. E. Howe, and P. D. Danley. 2014. Quantitative genetic analyses of male color pattern and female mate choice in a pair of cichlid fishes of Lake Malawi, East Africa. *PLOS One* 9:1–22.

Feulner, P. G. D., F. Kirschbaum, V. Mamonekene, V. Ketmaier, and R. Tiedemann. 2007. Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: Campylomormyrus): a combined molecular and morphological approach. *J. Evol. Biol.* 20:403–414.

Feulner, P. G. D., F. Kirschbaum, and R. Tiedemann. 2008. Adaptive radiation in the Congo River: An ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; Campylomormyrus). *J. Physiol. Paris* 102:340–346. Elsevier Ltd.

Freedman, E. G., J. Olyarchuk, M. A. Marchaterre, and A. H. Bass. 1989. A temporal analysis of testosterone-induced changes in electric organs and electric organ discharges of mormyrid fishes. *J. Neurobiol.* 20:619–634.

Gallant, J. R., M. E. Arnegard, J. P. Sullivan, B. A. Carlson, and C. D. Hopkins. 2011. Signal variation and its morphological correlates in *Paramormyrops kingsleyae* provide insight into the evolution of electrogenic signal diversity in mormyrid electric fish. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 197:799–817.

Gallant, J. R., L. L. Traeger, J. D. Volkening, H. Moffett, P.-H. Chen, C. D. Novina, G. N. Phillips, R. Anand, G. B. Wells, M. Pinch, R. Guth, G. A. Unguez, J. S. Albert, H. H. Zakon, M. P. Samanta, and M. R. Sussman. 2014. Genomic basis for the convergent evolution of electric organs. *Science* 344:1522–1525.

Givnish, T. J. 2015. Adaptive radiation versus “radiation” and “explosive diversification”: Why conceptual distinctions are fundamental to understanding evolution. *New Phytol.* 207:297–303.

Grant, B. R., and P. R. Grant. 2008. Fission and fusion of Darwin’s finches populations. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363:2821–2829.

Grant, R. B., and P. R. Grant. 2003. What Darwin’s Finches Can Teach Us about the Evolutionary Origin and Regulation of Biodiversity. *Bioscience* 53:965–975.

Haesler, M. P., and O. Seehausen. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: Implications for speciation. *Proc. Royal Soc. B* 272:237–245.

Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome Res.* 27:1004–1015.

Herfeld, S., and P. Moller. 1998. Effects of 17 α -methyltestosterone on sexually dimorphic characters in the weakly discharging electric fish, *Brienomyrus niger* (Gunther, 1866) (Mormyridae): Electric organ discharge, ventral body wall indentation, and anal-fin ray bone expansion. *Horm. Behav.* 34:303–319.

Hopkins, C. D. 1986. Behavior of Mormyridae. Pp. 527–576 in T. H. Bullock and W. Heiligenberg, eds. *Electroreception*. Wiley, New York.

Hopkins, C. D. 1999a. Design features for electric communication. *J. Exp. Biol.* 202:1217–1228. *Co Biol.*

Hopkins, C. D. 1981. On the diversity of electric signals in a community of mormyrid electric fish in West Africa. *Am. Zool.* 21:211–222. Oxford University Press.

Hopkins, C. D. 1999b. Signal evolution in electric communication. Pp. 461–491 in M. Hauser and M. Konishi, eds. *The design of animal communication*. MIT Press, Cambridge, MA.

Hopkins, C. D., and A. H. Bass. 1981. Temporal coding of species recognition signals in an electric fish. *Science* 212:85–87.

Hulsey, C. D. 2009. Cichlid genomics and phenotypic diversity in a comparative context. *Integr. Comp. Biol.* 49:618–629.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nat. Rev. Genet.* 5:288–298.

Kramer, B. 1974. Electric organ discharge interaction during interspecific agonistic behaviour in freely swimming mormyrid fish. *J. Comp. Physiol.* 93:203–235. Springer-Verlag.

Lamanna, F., F. Kirschbaum, I. Waurick, C. Dieterich, and R. Tiedemann. 2015. Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). *BMC Genomics* 16:668. BMC Genomics.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

Lamichhaney, S., F. Han, J. Berglund, C. Wang, M. S. Almén, M. T. Webster, B. R. Grant, P. R. Grant, and L. Andersson. 2016. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352:470–474. American Association for the Advancement of Science.

Lavoué, S., M. E. Arnegard, J. P. Sullivan, and C. D. Hopkins. 2008. Petrocephalus of Odzala offer insights into evolutionary patterns of signal diversification in the Mormyridae, a family of weakly electrogenic fishes from Africa. *J. Physiol. Paris* 102:322–339. Elsevier.

Lissmann, H. W., and K. E. Machin. 1958. The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J. Exp. Biol.* 35:451–486.

Magalhaes, I. S., and O. Seehausen. 2010. Genetics of male nuptial colour divergence between sympatric sister species of a Lake Victoria cichlid fish. *J. Evol. Biol.* 23:914–924. John Wiley & Sons, Ltd (10.1111).

Möhres, F. P. 1957. Elektrische Entladungen im Dienste der Revierabgrenzung bei Fischen. *Naturwissenschaften* 44:431–432. Springer-Verlag.

Mullen, S. P., T. C. Mendelson, C. Schal, and K. L. Shaw. 2007. Rapid evolution of cuticular hydrocarbons in a species radiation of acoustically diverse Hawaiian crickets (Gryllidae: Trigonidiinae: Laupala). *Evolution* 61:223–231.

Nagel, R., F. Kirschbaum, and R. Tiedemann. 2017. Electric organ discharge diversification in mormyrid weakly electric fish is associated with differential expression of voltage-gated ion channel genes. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 203:183–195. Springer Berlin Heidelberg.

O'Quin, C. T., A. C. Drilea, R. B. Roberts, and T. D. Kocher. 2012. A Small Number of Genes Underlie Male Pigmentation Traits in Lake Malawi Cichlid Fishes. *J. Exp. Zool. Part B Mol. Dev. Evol.* 318:199–208.

Peichel, C. L., and D. A. Marques. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 372.

Picq, S., J. Sperling, C. J. Cheng, B. A. Carlson, and J. R. Gallant. 2020. Genetic drift does not sufficiently explain patterns of electric signal variation among populations of the mormyrid electric fish *Paramormyrops kingsleyae*. *Evolution* 1–25.

Podos, J. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* 409:185–188.

- Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4:1–8.
- Seehausen, O. 2006. African cichlid fish: A model system in adaptive radiation research. *Proc. Royal Soc. B* 273:1987–1998.
- Servedio, M. R., and J. W. Boughman. 2017. The Role of Sexual Selection in Local Adaptation and Speciation. *Annu. Rev. Ecol. Evol. Syst.* 48:85–109.
- Servedio, M. R., G. S. Van Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: “magic” but not rare? Elsevier Current Trends.
- Shaw, K. L. 2000. Further acoustic diversity in Hawaiian forests: two new species of Hawaiian cricket (Orthoptera: Gryllidae: Trigonidiinae: Laupala). *Zool. J. Linn. Soc.* 129:73–91.
- Shaw, K. L., and S. C. Lesnick. 2009. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl. Acad. Sci. U.S.A.* 106:9737–9742. National Academy of Sciences.
- Sullivan, J. P., S. Lavoué, M. E. Arnegard, and C. D. Hopkins. 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* 58:825–41.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2002. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution* 56:597–616. Society for the Study of Evolution.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2000. Molecular systematics of the African electric fishes (Mormyroidea: teleostei) and a model for the evolution of their electric organs. *J. Exp. Biol.* 203:665–683.
- Szabo, T. 1961. Les Organes Electriques des Mormyrides. Pp. 20–24 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, New York.
- Traeger, L. L., J. D. Volkening, H. Moffett, J. R. Gallant, P.-H. Chen, C. D. Novina, G. N. Phillips, R. Anand, G. B. Wells, M. Pinch, R. Güth, G. A. Unguez, J. S. Albert, H. Zakon, M. R. Sussman, and M. P. Samanta. 2015. Unique patterns of transcript and miRNA expression in the South American strong voltage electric eel (*Electrophorus electricus*). *BMC Genomics* 16:243.
- von der Emde, G., M. Amey, J. Engelmann, S. Fetz, C. Folde, M. Hollmann, M. Metzen, and R. Pusch. 2008. Active electrolocation in *Gnathonemus petersii*: Behaviour, sensory performance, and receptor systems. *J. Physiol. Paris* 102:279–290. Elsevier Ltd.
- Wiley, C., C. K. Ellison, and K. L. Shaw. 2012. Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. *Proc. Royal Soc. B* 279:1203–1209. The Royal Society.

CHAPTER 1: THE TRANSCRIPTIONAL CORRELATES OF DIVERGENT ELECTRIC ORGAN DISCHARGES IN
PARAMORMYROPS ELECTRIC FISH¹

Introduction

Understanding the genomic basis of phenotypic diversity is a major goal of evolutionary biology (Butlin et al. 2012). Adaptive radiations and explosive diversification of species (Givnish 2015) are frequently characterized by interspecific phenotypic differences in divergence of few, hypervariable phenotypic traits (Shaw 2000; Grant and Grant 2003; Kocher 2004; Mullen et al. 2007). Such systems offer exceptional advantages to study the genomic bases of phenotypic diversity: they can provide replication under a controlled phylogenetic framework (Seehausen 2006), and couple ample phenotypic differentiation with relatively “clean” genomic signals between recently diverged species (Hulsey 2009). Study of the genomic mechanisms underlying hypervariable phenotypic traits has identified, in some cases relatively simple genetic architectures (Magalhaes and Seehausen 2010; O’Quin et al. 2012; Lamichhaney et al. 2015, 2016; Han et al. 2017). More often, the genetic architecture underlying such traits can be complex and polygenic (Shaw and Lesnick 2009; Wiley et al. 2012; Ding et al. 2014; Blankers et al. 2018). It has long been recognized that changes in gene expression can affect phenotypic differences between species (King and Wilson 1975), and RNA-seq based approaches have greatly facilitated the study of this relationship (Alvarez et al. 2015). A growing number of studies have examined differences in gene expression in phenotypic evolution (e.g., (Jeukens et al. 2010; Brawand et al. 2011; Ferreira et al. 2013; Alvarez et al. 2015; Abolins-Abols et al. 2018; Giorello et al. 2018; Härer et al. 2018; Nigenda-Morales et al. 2018; Young et al. 2019)). While these studies do not implicate mutational causes, analysis of differential gene expression (DGE) can be a useful approach in examining the genomic basis of divergent phenotypes.

¹ This work has been published as: Losilla, M., D. M. Luecke, and J. R. Gallant. 2020. The transcriptional correlates of divergent electric organ discharges in *Paramormyrops* electric fish. *BMC Evol. Biol.* 20:6. Please refer to this publication for the Additional Files.

African weakly electric fish (Teleostei: Mormyridae) are among the most rapidly speciating groups of ray-finned fishes (Carlson et al. 2011; Rabosky et al. 2013). This is partly due to the diversification of the genus *Paramormyrops* (Sullivan et al. 2002, 2004) in the watersheds of West-Central Africa, where more than 20 estimated species (Lavoué et al. 2008) have evolved within the last 0.5-2 million years (Sullivan et al. 2002). Extensive evidence has demonstrated that electric organ discharges (EODs) exhibit little intraspecific variation, yet differ substantially among mormyrid species (Hopkins 1981; Bass 1986; Arnegard et al. 2005). This pattern is particularly evident in *Paramormyrops* (Sullivan et al. 2000, 2002), in which EOD waveforms evolve much faster than morphology, size, and trophic ecology (Arnegard et al. 2010a).

Mormyrid EODs are a behavior with a dual role in electrolocation (Lissmann and Machin 1958; von der Emde et al. 2008) and intraspecific communication (Möhres 1957; Kramer 1974). EOD waveforms vary between species principally in terms of their complexity, polarity, and duration (Hopkins 1999b; Sullivan et al. 2002), and all three dimensions of variation are evident among *Paramormyrops* (Fig. 1.1). Furthermore, recent discoveries of intraspecific polymorphism in EOD waveform in *P. kingsleyae* (Gallant et al. 2011) and polarity among *P. sp. 'magnostipes'* (Arnegard et al. 2005) present a unique opportunity to study the genomic basis of phenotypic traits within a rapidly diverging species group.

EODs have a well-understood morphological (Fig. 1.1) and neurophysiological basis (Carlson 2002; Caputi et al. 2005). EODs are generated by specialized cells (electrocytes) that constitute the electric organ (EO), located in the caudal peduncle (Hopkins 1986). Mormyrid EOs are comprised of 80-360 electrocytes (Bass 1986), and an individual EOD is produced when the electrocytes discharge synchronously. EODs are multiphasic because they result from action potentials produced by two excitable membranes: the two large phases of the EOD, called P1 and P2, are produced by spikes generated by the posterior and anterior electrocyte faces, respectively (Bennett and Grundfest 1961).

There is a relationship between EODs of longer duration and increased surface membrane area (Bennett 1971), likely mediated at least in part by an increase in membrane capacitance (Bass et al. 1986; Bass and Volman 1987). The duration of EODs is highly variable within mormyrids-- some EODs are extremely long (>15 ms) and others are very brief (0.2 ms) (Lavoué et al. 2008).

Within the Mormyridae, triphasic EODs evolved early from biphasic EODs; however, there have been multiple parallel reversions to biphasic EODs across mormyrids and within *Paramormyrops* (Sullivan et al. 2000; Gallant et al. 2011). Triphasic (P0-present) EODs are produced by electrocytes that are innervated on the anterior face and have penetrating stalks (*Pa*, P-type), whereas biphasic (P0-absent) EODs are produced by electrocytes innervated on the posterior face and lack penetrating stalks (*NPp*, N-type) (for more details see (Bennett and Grundfest 1961; Szabo 1961; Bennett 1971; Hopkins 1999b; Gallant et al. 2011; Carlson and Gallant 2013)). We refer to triphasic EODs as more ‘complex’ than biphasic EODs. In some cases, triphasic EODs display an unusually large P0 phase, which gives the appearance of an ‘inverted’ polarity. This is exemplified by the type I EODs of *P. sp. ‘magnostipes’* (Fig. 1.1) (Arnegard et al. 2005). The number (Bennett and Grundfest 1961) and diameter (Bass 1986; Gallant et al. 2011) of stalk penetrations are positively correlated with the magnitude of P0. We refer to individuals with large penetrations as ‘inverted’ polarity and individuals with small penetrations as ‘normal’ polarity.

Recent studies in mormyrids (Zakon et al. 2006; Arnegard et al. 2010b; Paul et al. 2016; Nagel et al. 2017; Swapna et al. 2018) have adopted a candidate gene approach to examine the molecular basis of variation in EOD duration on macroevolutionary scales, implicating voltage gated sodium channels (e.g. *scn4aa*) and potassium channels (e.g. *kcna7a*) as key targets of selection during EOD evolution. Beyond this recent attention to ion channels, several studies have described the importance of structural differences between EOs as an important component of EOD variation (Bennett 1971; Bass et al. 1986; Gallant et al. 2011). In this study, we took a transcriptome-wide approach to characterizing the

molecular basis of electric signal diversity in *Paramormyrops* species divergent for EOD complexity, duration and polarity. We used RNA-sequencing to comprehensively examine DGE in the adult EOs of five *Paramormyrops* operational taxonomic units (OTUs), leveraging a recently sequenced and annotated genome assembly from the species *P. kingsleyae* (N-type) (Gallant et al. 2017), and identify gene expression correlates of each of the three main EOD waveform features of electric signal diversity in *Paramormyrops*. Our results emphasize genes that influence the shape and structure of the electrocyte cytoskeleton, membrane and extracellular matrix (ECM) to exhibit predictable differences between *Paramormyrops* species with divergent EOD phenotypes.

Methods

Sample collection

We captured 11 *Paramormyrops* individuals from Gabon, West Central Africa in 2009: five *P. kingsleyae* (n=3 N-type and n=2 P-type), four *P. sp.* ‘magnostipes’ (n=2 Type I and n=2 Type II), and two *P. sp.* ‘SN3’. Within 1-12 hours of capture, individual specimens were euthanized by overdose with MS-222. The caudal peduncle was excised and skinned, and immediately immersed in RNA-later for 24h at 4°C, before being transferred to -20°C for long-term storage. As two of these species (*P. sp.* ‘magnostipes’, *P. sp.* ‘SN3’) are presently undescribed, we note that these specimens were identified by their EOD waveform, head morphology and collecting locality (Sullivan et al. 2002, 2004; Arnegard et al. 2005, 2006). All specimens, including vouchers materials, are deposited in the Cornell University Museum of Vertebrates. Collection information and the phenotypes per EOD feature of each sample are detailed in Table 1.1.

RNA extraction, cDNA library preparation and Illumina Sequencing

Total RNA was extracted from EOs using RNA-easy Kit (Qiagen, Inc) after homogenization with a bead-beater (Biospec, Inc.) in homogenization buffer. mRNA was isolated from total RNA using a NEBNext mRNA Isolation Kit (New England Biolabs, Inc.). Libraries for RNA-seq were prepared using the NEBNext mRNA Sample Prep Master Mix Set, following manufacturer’s instructions. Final libraries after

size selection ranged from 250-367 bp. Libraries were pooled and sequenced by the Cornell University Biotechnology Resource Center Genomics Core on an Illumina HiSeq 2000 in a 2x100bp paired end format. Raw sequence reads were deposited in the NCBI SRA (Additional file 1).

Read processing and data exploration

FastQC v0.11.3 (Babraham Bioinformatics) was used to manually inspect raw and processed reads. We used Trimmomatic v.0.32 (Bolger et al. 2014) to remove library adaptors, low quality reads, and filter small reads; following the suggested settings of (MacManes 2014): 2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25. After trimming, reads from each specimen were aligned to the predicted transcripts of the NCBI-annotated (Release 100) *P. kingsleyae* (N-type) genome (Gallant et al. 2017) using bowtie 2 v2.3.4.1 (Langmead and Salzberg 2012). Expression quantification was estimated at the gene level using RSEM v1.3.0 (Li and Dewey 2011), followed by exploration of the data with a gene expression correlation matrix based on Euclidean distances and Pearson's correlation coefficient (for genes with read counts >10, Trinity's default parameters). All these steps were executed using scripts included with Trinity v2.6.6 (Grabherr et al. 2011; Haas et al. 2013).

Data Analysis

We began by examining DGE between all possible pairwise comparisons of OTUs (n=10, Table 1.2) using edgeR v3.20.9 (Robinson and Oshlack 2010) through a script provided with Trinity (MA plots are provided in Additional File 8). We restricted our consideration of genes to those where CPM-transformed counts were > 1 in at least two samples for each comparison (edgeR default parameters). We modified this to use the function `estimateDisp()` instead of the functions `estimateCommonDisp()` and `estimateTagwiseDisp()`. For each comparison, we conservatively considered genes to be differentially expressed with a minimum fold change of 4 and p-value of 0.001 after FDR correction. We compiled a non-redundant list of genes that were differentially expressed in at least one comparison based on these criteria (Fig. 1.2, Set A).

For each of the DEGs in Set A we used TMM normalized values to compare gene expression between groups of OTUs with alternative EOD waveform phenotypes (i.e. *long duration EOD* vs. *short duration EOD*, *biphasic* vs. *triphasic* and *small penetrations* vs. *large penetrations*, see Table 1.1. Note that waveform polarity phenotypes only apply to triphasic individuals). For each of the three phenotype pairs, we calculated the mean and standard deviation for TMM values within each grouping, then extracted the genes that had (1) expression values more than four times greater in one phenotype than the other and (2) a difference in mean expression greater than either within-group standard deviation. This resulted in six lists of upregulated genes, one for each EOD feature across all OTUs and samples (Fig. 1.2, Set B).

In order to assess enrichment of particular gene pathways, biological functions, and cellular locations using a controlled vocabulary, we performed GO (Ashburner et al. 2000; The Gene Ontology Consortium 2019) enrichment tests on every list of upregulated genes from (1) the ten pairwise comparisons (n=20, two per comparison), (2) Set B (n=6), and (3) Set C (n=6), for each of the three ontology domains: Biological Process, Cellular Component, and Molecular Function. First, we identified homologous proteins predicted from the *P. kingsleyae* (N-type) reference genome and those predicted from *Danio rerio* (GRCz11) by blastp (BLAST+ v2.6, (Camacho et al. 2009)). For each protein, the top hit (e-value $\leq 1e-10$) was used for annotation. Next, we used mygene v1.14.0 (Wu et al. 2013; Xin et al. 2016) to match the *D. rerio* proteins to *D. rerio* genes and extract their GO annotations (zebrafish Zv9). This resulted in GO annotations for each of the three ontology domains for *P. kingsleyae* (N-type) genes. Finally, we carried out the GO enrichment tests using topGO v2.30.1 (Alexa et al. 2006) and the following parameters: nodeSize = 10, statistic = fisher, algorithm = weight01, p-value ≤ 0.02 . The ‘universe’ for each enrichment test on gene lists from the pairwise comparisons was all the genes deemed expressed in the respective comparison, whereas the non-redundant list of genes in these ten ‘universes’ was the ‘universe’ for all enrichment tests on the gene lists from Sets B and C.

Interpretation of lists of genes from Set A and Set B each suffered limitations for the overall goals of this analysis, which is to identify the DEGs most strongly associated with each waveform feature (duration, complexity, and polarity). The ten comparisons made to construct Set A were not equally informative for two primary reasons: (1) the OTUs in this analysis vary in terms of their phylogenetic relatedness (see (Sullivan et al. 2002, 2004)) and (2) several OTU comparisons varied in more than one waveform characteristic (Table 1.2). As such, we elected to focus on the most informative comparison for each EOD feature: the comparison that contrasted only the given feature and that minimized phylogenetic distance between OTUs. Of the ten pairwise comparisons, we classified three as the most informative comparisons, one per EOD feature (Table 1.2). The six lists of significantly upregulated genes (fold change > 4, FDR-corrected p-value < 0.001) from these three pairwise OTU comparisons constitute Set A'.

Comparisons in Set A', however, lack biological replication. In contrast, interpretations of Set B were potentially limited in that many of the OTUs in this analysis differed in more than one EOD feature. To circumvent the limitations of Sets A' and B within the limits of our study design, we constructed a third set (Set C). Set C is defined as the intersection of the upregulated genes from Sets A' and B, for each phenotype. Since there were six phenotypes in our study, Set C encompasses six lists of upregulated genes and their respective enriched GO terms (Fig. 1.2). Therefore, Set C represents the DEGs that are (1) statistically supported as differentially expressed between closely-related OTUs that vary in a single waveform characteristic, and (2) are consistently differentially expressed among all OTUs that share that waveform feature. We focus our attention on Set C: We retrieved GO term definitions from QuickGO (Binns et al. 2009) and descriptions of gene function of the functional annotations from UniProt (The UniProt Consortium 2019); and to facilitate the discussion, we classified the more interesting genes in Set C into "general" functional classes, or themes.

Results

Overall Results

We extracted and sequenced mRNA from EOs of 11 wild caught *Paramormyrops* samples from five OTUs (Table 1.1). Overall alignment rates of the processed reads to the *Paramormyrops kingsleyae* reference transcriptome ranged from 28-74% (>375 million sequenced reads in total, 50% aligned), with no clear differences among OTUs (Additional file 1). On inspection, we concluded that these rates are a consequence of the presence of overrepresented sequences from rRNA, mtDNA and bacterial contamination in the RNA-seq reads.

We explored the data with a heatmap of pairwise correlations of gene expression for 24,960 genes across all 11 samples (Fig. 1.3), and carried out all possible pairwise DGE comparisons of OTUs (n=10, Table 1.2). These ten comparisons detected a range of 16,420-19,273 expressed genes. Intersection of these lists resulted in a non-redundant list of 20,197 genes expressed in EO across all DGE comparisons. We found that 3,274 (16%) were differentially expressed in at least one comparison, and expression patterns across all OTUs were highly correlated (Pearson's $r > 0.89$, Fig. 1.3). Despite this, correlation values were higher within recognized OTUs, except for the *P. sp.* 'magnostipes type II' 6768 sample (Fig. 1.3). Thus, we did not use *P. sp.* 'magnostipes type I' vs *P. sp.* 'magnostipes type II' as the informative comparison for waveform polarity in Set A' (see methods).

Set A: Differential Expression Analysis

We found between 489-1542 differentially expressed genes (DEGs; fold change > 4, FDR-corrected p-value < 0.001) (50-128 enriched Gene Ontology (GO) terms) in every pairwise comparison of OTUs except *P. sp.* 'magnostipes type I' vs *P. sp.* 'magnostipes type II', which had only nine DEGs with seven enriched GO terms (Table 1.2). Additional file 2 provides a tabular list of DEGs for each comparison, and Additional file 3 provides a tabular list of enriched GO terms for each comparison. We call Set A the non-redundant list of 3,274 genes that were differentially expressed in at least one DGE comparison (Fig. 1.2, Set A).

We chose the phylogenetically most informative comparisons to construct Set A', which are indicated in Table 1.2. We found: 507 DEG and 69 enriched GO terms comparing *P. kingsleyae* (N-type) vs *P. sp.* 'SN3' (EOD duration); 1322 DEG and 77 enriched GO terms comparing *P. kingsleyae* (P-type) vs *P. sp.* 'magnostipes type I' (waveform polarity); and 530 DEG and 75 enriched GO terms comparing *P. kingsleyae* (N-type) vs *P. kingsleyae* (P-type) (waveform complexity).

Set B: Expression Based Clustering

For each EOD feature (n=3), we grouped OTUs by phenotype (Table 1.1), and calculated normalized expression values for Set A genes. From these, we constructed Set B by selecting genes that (1) exhibit greater than four-fold difference in the average expression levels between phenotypes of each EOD feature, and (2) have within-phenotype standard deviations less than the difference between phenotype-mean expression. For the phenotypes of waveform duration, we identified a combined total of 309 DEG and 43 enriched GO terms, for waveform polarity we found 169 DEG and 14 enriched GO terms, and for waveform complexity the totals were 413 DEG and 38 enriched GO terms. Additional file 4 lists the identities of these DEG and Additional file 5 lists their enriched GO terms for all three GO ontologies.

Set C: Intersection of Phylogenetically Informative Comparisons and Expression Based Clustering

We were motivated to obtain the DEGs and enriched GO terms that were most likely to be associated with divergent EOD phenotypes. To obtain this list, we constructed Set C, which is the intersection of Set A' and Set B (Fig. 1.2, see Methods). The expression profiles of the Set C genes for each EOD feature, along with the enriched GO terms for Biological Process and Cellular Component, are shown in Figs. 4-6.

Contrast of waveform duration identified 183 DEG and 39 enriched GO terms. 140 of the DEG were upregulated in the short EOD phenotype (Fig. 1.4 purple lines), and 43 genes were upregulated in samples with long EODs (Fig. 1.4 yellow lines). Contrast of waveform polarity identified 154 DEG and 17 enriched GO terms. We found 99 upregulated genes in individuals with small penetrations (Fig. 1.5 red

lines), and 55 upregulated genes in individuals with large penetrations (Fig. 1.5 grey lines). Finally, contrast of waveform complexity identified 145 DEG and 20 enriched GO terms. We detected 110 upregulated genes in individuals with biphasic EODs (Fig. 1.6 blue lines), and 35 upregulated genes in individuals with triphasic EODs (Fig. 1.6 orange lines). These results are further detailed in Table 1.3. All genes in Set C are listed in Additional file 6 and their enriched GO terms are listed in Additional file 7.

Discussion

It has long been recognized that changes in gene expression can affect phenotypic differences between species (King and Wilson 1975), and RNA-seq has facilitated the study of this relationship (Alvarez et al. 2015). The goal of this study was to determine DEGs associated with divergent EOD features within *Paramormyrops*. Expression patterns across all OTUs were highly correlated (Pearson's $r > 0.89$, Fig. 1.3) and we detected differential expression of only 3,274 (16%) genes between any two OTUs. Thus, a major finding of this study is that EO gene expression is overall quite similar across *Paramormyrops* species with divergent EODs, and relatively few genes are associated with phenotypic differences in EOD waveform between OTUs. We find this notable given observations of generally high levels of genetic distances between geographically proximate populations of the same *Paramormyrops* species (Arnegard et al. 2005; Picq et al. 2020).

Despite the relatively small number of DEGs compared to the total number of genes expressed in the EO, we constructed our analysis to extract genes that were highly associated with particular phenotypes. Set A' represents a formal statistical test that contrasted OTUs. Each comparison contrasted samples from OTUs that were divergent in only one EOD feature, while minimizing phylogenetic distance. The drawback of this approach is the observed differences may not reflect a general pattern across multiple OTUs, instead resulting from OTU-specific changes or confounding variables such as collection sites. Set B took the opposite approach, using information from all possible biological replicates to identify consistent gene expression patterns between EOD phenotypes, at the

expense of formal statistical support and the introduction of confounding phylogenetic relationships and phenotypic heterogeneity. To balance these drawbacks, we constructed Set C, which represents genes and GO terms that are differentially expressed/enriched between closely related OTUs divergent in only one phenotypic character and that are also consistently differentially expressed/enriched among representatives with similar EOD phenotypes. As such, we focus our discussion on the results of Set C. We classified the genes in Set C into “general” functional classes, or themes; and focus our attention on the ones that relate to the known morphological underpinnings of waveform duration (Table 1.4), polarity (Table 1.5), and complexity (Table 1.6). These functional classes were genes related to the ECM, cation homeostasis, lipid metabolism, and cytoskeletal and sarcomeric genes.

Waveform Duration

Several researchers have implicated the role of ion channels in the evolution of duration changes in mormyrid signals (Zakon et al. 2006; Arnegard et al. 2010b; Paul et al. 2016; Nagel et al. 2017; Swapna et al. 2018). We did not find evidence of large changes in expression of sodium channels between short-duration *P. sp. ‘SN3’* and other *Paramormyrops* species; but we detected upregulation in short EOD samples of a voltage insensitive, outwardly rectifying potassium channel (*potassium channel subfamily K member 5-like*) and of a regulatory subunit of a Shal-type voltage-gated potassium channel (*Kv channel-interacting protein 4*). Additionally, individuals with short EODs upregulate two calcium-binding proteins: *parvalbumin-2*, and *parvalbumin-2-like*. Parvalbumins are highly expressed in skeletal muscle where they sequester calcium after contraction, thus facilitating relaxation. Frequently, muscles with fast relaxation rates express higher levels of parvalbumins (Wilwert et al. 2006). The upregulated parvalbumin genes we detected may somehow be related to shorter EODs by sequestering calcium at a faster rate, which could affect action potentials directly or indirectly through calcium-activated ion channels.

Previous studies have demonstrated that changes in EOD duration result from changes in electrocyte ultrastructure. The two major phases of the EOD waveform are caused by action potentials

generated by the anterior and posterior faces (Bennett and Grundfest 1961). Bennett demonstrated a relationship between EOD duration and increased surface membrane area (Bennett 1971), and Bass et al. showed that differences in surface area are more readily noticeable on the anterior face (Bass et al. 1986). Membrane surface area is increased by folding the electrocyte membrane into papillae and other tube-like invaginations (Schwartz et al. 1975). Testosterone can induce increases in EOD duration in several mormyrids (Bass and Hopkins 1983, 1985; Bass et al. 1986; Bass and Volman 1987), and it also increases membrane surface area, either particularly on the anterior face (Bass et al. 1986) or on both anterior and posterior faces (Freedman et al. 1989). A larger surface area may increase the capacitance of the membrane, thus delaying spike initiation (Bass et al. 1986; Bass and Volman 1987). Consequently, genes involved in the synthesis of membranes could influence EOD duration.

We found the most prominent differences in gene expression between the EOD duration phenotypes in genes that code for cytoskeletal, sarcomeric, and lipid metabolism proteins (Table 1.4). We emphasize the last group: no lipid metabolism genes were upregulated in individuals with long EODs, whereas samples with short EODs upregulated *protein EFR3 homolog B-like* (a regulator of phosphatidylinositol 4-phosphate synthesis), *retinoic acid receptor responder protein 3-like* and *HRAS-like suppressor 3* (these two catalyze hydrolysis of phosphatidylcholines and phosphatidylethanolamines), *fatty aldehyde dehydrogenase-like* (fatty acid metabolism), *PTB domain-containing engulfment adapter protein 1-like* (modulates cellular glycosphingolipid and cholesterol transport), *phosphatidylinositol 3-kinase regulatory subunit gamma-like*, (PI3K, which phosphorylates phosphatidylinositol), and *proto-oncogene c-Fos-like* (can activate phospholipid synthesis), and showed enrichment of the GO term ‘phosphatidylinositol phosphorylation.’ We hypothesize that these genes are involved in the surface proliferation of the electrocytes membranes.

Additionally, each mormyrid electrocyte stands embedded in a gelatinous mucopolysaccharide matrix (the ECM) separated from neighboring electrocytes by connective tissue septa (Fig. 1.1) (Bass

1986), and the membrane surface invaginations are coated by the same ECM that surrounds the electrocytes (Schwartz et al. 1975; Bass et al. 1986). Hence, differences in surface invaginations could also be reflected in differences in the expression of genes whose products interact with the ECM. In individuals with long EODs, we found upregulated the genes *matrilin 2* (involved in matrix assembly) and *thrombospondin-4-like* (mediates cell-to-matrix interactions), and enriched the GO term ‘extracellular space’; whereas those with short EODs upregulated *collagenase 3-like* (plays a role in the degradation of ECM proteins) and displayed enrichment of the GO terms ‘extracellular region’ and ‘external side of plasma membrane’.

Two previous studies focused on DGE between EOs in another mormyrid species adaptive radiation/explosive diversification (genus *Campylomormyrus*, whose EO anatomy closely resembles that of *Paramormyrops*). Both focused on comparisons between species with biphasic EODs but different waveform duration phenotypes: *C. tshokwe* (long duration) and *C. compressirostris* (short duration). The first study performed a candidate gene approach to quantify the expression patterns of 18 sodium and potassium homeostasis genes between the EOs of the two species (Nagel et al. 2017), whereas Lamanna et al. (Lamanna et al. 2015) used RNA-seq to simultaneously compare gene expression between EOs of these species. While we did not observe differences in expression of any of the potassium channels reported by Nagel et al., we note that Lamanna et al. reported differential expression of metabolic pathways related genes, particularly fatty acid metabolism, and ion transport and neuronal function (referred to in their text as cross-species analysis (EO) subclusters 2 and 4). While we found no overlap in the identities of any specific genes in our study, we note that our analysis also detected differential expression of lipid metabolism related genes when comparing EODs of different duration.

Overall, our results identify genes that may affect EOD duration through membrane rearrangements, which could be coupled with changes in the interaction with the ECM and the expression of cytoskeletal and sarcomeric genes. Since this waveform feature is modulated by

testosterone, this androgen could facilitate the study of these suggested genetic underpinnings under more rigorously controlled circumstances.

Waveform Polarity

The number (Bennett and Grundfest 1961) and diameter (Bass 1986; Gallant et al. 2011) of stalk penetrations are positively correlated with the magnitude of P0. This phenomenon is exemplified by *P.* sp. ‘magnostipes type I’, which has the largest P0 in the OTUs examined in this study, giving the EOD the appearance that it ‘inverted’ relative to other EODs. This OTU has numerous, large diameter penetrations, whereas *P. kingsleyae* (P-type) has relatively fewer, small diameter penetrations (Fig. 1.1). These large structural differences may influence the electrocyte’s connection with the surrounding ECM, and our results support this: the phenotypes of waveform polarity exhibited differences in the expression of genes that interact with the extracellular space. We found no such genes upregulated in individuals with large penetrations, whereas in samples with small penetrations we detected the enriched GO terms: ‘extracellular matrix structural constituent’ and ‘basement membrane,’ and the upregulated genes: *collagen alpha-1(V) chain-like*, *collagen type IX alpha 2 chain*, *collagen alpha-4(VI) chain-like*, *epiphycan-like* (may play a role in cartilage matrix organization), *cell migration-inducing and hyaluronan-binding protein-like* (mediates depolymerization of hyaluronic acid), *inter-alpha-trypsin inhibitor heavy chain H5-like* (although this gene is little studied, inter-alpha-trypsin inhibitors usually interact with hyaluronan), and *thrombospondin 2* (mediates cell-to-matrix interactions).

OTUs with small penetrations also exhibited higher expression of genes related to cytoskeletal, sarcomeric, and lipid metabolism proteins than do individuals with large penetrations (Table 1.5). This includes the genes *myosin light chain 3-like*, *desmin-like*, *PH* and *SEC7 domain-containing protein 1-like* (induces cytoskeletal remodeling), *CDC42 effector protein 3* (probably involved in the organization of the actin cytoskeleton), *protein phosphatase 1 regulatory subunit 12B-like* (regulates myosin phosphatase activity), *protein kinase C alpha type-like* (phosphorylates cardiac troponin T), and *calponin-1-like* (modulates smooth muscle contraction). Samples with large penetrations showed upregulation of the

genes *myosin light chain 4-like* (regulatory light chain of myosin), *leucine rich repeat containing 10*, (may play important roles in cardiac development and/or function), and *Wiskott-Aldrich syndrome protein family member 3-like* (regulation of cell morphology and cytoskeletal organization).

We hypothesize that the differences in the number and diameter of penetrations that drive variation in EOD waveform polarity require changes to the electrocyte's cytoskeletal and membrane properties. These arrangements may be necessary for the electrocytes body to adjust to the increased volume displacements imposed by larger penetrations; or alternatively, they may be a prerequisite for penetrating stalks to enlarge. Our observations support and elaborate on the hypothesis that sarcomeric proteins (which are non-contractile in mormyrid electric organs) may function as a means of cytoskeletal support and structural integrity in mormyrid electrocytes (Gallant et al. 2012).

Waveform Complexity

Waveform complexity refers to the number of phases present in an EOD, and mormyrid EODs vary in the presence of a small head negative phase (P0). The presence or absence of P0 in the EOD depends on the anatomical configuration of the electrocytes: P0-present (or triphasic) EODs are produced by electrocytes that are innervated on the anterior face and have penetrating stalks (Pa), whereas P0-absent (or biphasic) EODs are produced by electrocytes innervated on the posterior face and lack penetrating stalks (NPp) (Bennett and Grundfest 1961; Szabo 1961; Bennett 1971; Hopkins 1999b; Gallant et al. 2011; Carlson and Gallant 2013). Developmental studies of the adult EO suggest that Pa electrocytes go through a NPp stage before developing penetrations (Szabo 1960; Denizot et al. 1982). This motivated the hypothesis that penetrations develop by the migration of the posteriorly innervated stalk system (NPp stage) through the edge of the electrocyte, and that the interruption of this migration represents a mechanism for Pa-to-NPp reversals (Alves-Gomes and Hopkins 1997; Hopkins 1999a).

Our data indicates several DEGs that implicate specific cytoskeletal and ECM reorganizations between triphasic and biphasic EODs (Table 1.6). We observed differential expression of several genes

associated with the polymerization of F-actin. In triphasic individuals, we observe upregulation of the gene *capping protein regulator and myosin 1 linker 3 (carmil3)*; although this gene is little studied, its paralog *carmil2* enhances F-actin polymerization. In contrast, the biphasic phenotype upregulated the genes *protein-methionine sulfoxide oxidase mical2b-like* (promotes F-actin depolymerization), *transmembrane protein 47-like* (may regulate F-actin polymerization), *5'-AMP-activated protein kinase subunit gamma-2-like* (could remodel the actin cytoskeleton), *FYVE*, *RhoGEF* and *PH domain-containing protein 4-like* (regulates the actin cytoskeleton), *leiomodin 2* (promotes actin polymerization, and required for normal sarcomere organization in the heart) and *family with sequence similarity 110 member C* (may play a role in microtubule organization). Thus, biphasic and triphasic EODs display several DEG, with potentially diverging outcomes, that influence the cellular internal structure.

We hypothesize that electrocytes with penetrating stalks (which produce triphasic EODs) require cytoskeletal arrangements to produce penetrations, perhaps related to increasing F-actin, to maintain their structural integrity. Similar to what we propose under waveform polarity, these arrangements may be necessary for the electrocyte body to adjust to the penetrations; or alternatively, they may be a prerequisite for penetrations to occur.

We also observed differential expression in proteins expressed in the ECM. In biphasic OTUs, we found the GO term ‘mucopolysaccharide metabolic process’ to be enriched, and two upregulated copies of the gene *inter-alpha-trypsin inhibitor heavy chain 3*, which may act as a binding protein between hyaluronan and other ECM proteins. In triphasic individuals, we found the upregulated genes *epiphycan-like*, which may play a role in cartilage matrix organization, and *ependymin-like* (ortholog to the zebrafish endymin-like gene *epdl2*).

Two endymin-like genes are among the most differentially expressed genes in the Set A' comparison for waveform complexity *P. kingsleyae* (N-type) vs *P. kingsleyae* (P-type) (500-fold more highly expressed in *P. kingsleyae* (P-type), Additional file 2). Although expressed in many tissues and

with little amino acid similarity, all ependymin-related proteins are secretory, calcium-binding glycoproteins that can undergo conformational changes and associate with collagen in the ECM. They have been involved in regeneration, nerve growth, cell contact, adhesion and migration processes (Ganss and Hoffmann 2009). We hypothesize that ependymin-related proteins, and potentially some of the other ECM proteins highly expressed in triphasic individuals, are part of the “fibrillar substance” that lies between the stalk and the electrocyte body in individuals with penetrating electrocytes (Bass et al. 1986). Notably, the *P. kingsleyae* genome assembly, which is based on a biphasic individual, contains three paralogs of *epdl2*, whereas the osteoglossiform *Scleropages formosus* only has one, suggesting the intriguing possibility that this gene may have been duplicated in *Paramormyrops* or in mormyrids. Ependymin-related paralogs have been proposed as suitable targets to experimentally test gene subfunctionalization (Suárez-Castillo and García-Arrarás 2007).

Altogether, our results for EOD waveform complexity suggest that the conformation of the cytoskeleton and the expression of proteins secreted to the ECM are important elements of the stalk penetrations, which generate triphasic EODs.

Conclusions

The widespread differential expression within *Paramormyrops* of calcium-related genes (Additional file 6) emphasizes a much-needed area of future research. Calcium is known to be necessary for the proper electrocyte repolarization in some gymnotiform species (Bartels 1971), but it may not be as important in others (Ferrari and Zakon 1993). Few studies have addressed calcium physiology in mormyrids: calcium-related proteins have been reported as differentially expressed in EO vs skeletal muscle in *Campylomormyrus* (Lamanna et al. 2015) and in *Brienomyrus brachyistius* (Gallant et al. 2012). As electrocytes do not contract, calcium may act in electrocytes as an important second messenger or cofactor, participate in interactions with the ECM, and/or to contribute to the electrocyte’s electrical properties through interaction with voltage gated ion channels.

A second notable pattern in our results is the unusual degree to which mormyrid electrocytes retain expression of some sarcomeric genes, which has been noted in several studies (Gallant et al. 2012, 2014, 2017; Lamanna et al. 2014, 2015). The role these proteins serve in electrocytes is presently unknown; however, our results indicate that some are highly differentially expressed between *Parmormyrops* with different EOD waveforms (Table 1.4, Table 1.5, & Table 1.6). This strongly suggests that sarcomeric proteins could play an important role in the conformational changes required to develop and sustain penetrations.

Finally, the biochemical composition and function of the ECM in electrocytes is poorly understood. Our analysis identifies differential expression in ECM-related genes across the genus *Paramormyrops*, associated with each of the three EOD features studied. At least four of these genes (*cell migration-inducing and hyaluronan-binding protein-like*, *inter-alpha-trypsin inhibitor heavy chain H5-like* and two copies of *inter-alpha-trypsin inhibitor heavy chain 3*), distributed across two EOD features, interact with hyaluronan. Hyaluronan is a type of mucopolysaccharide and a major component of some soft tissues and fluids (Fraser et al. 1997). Therefore, we propose that hyaluronan is an important constituent of the ECM in mormyrid fish. In addition, the electrocyte-ECM interactions should be an important area of future investigation, as they are likely to influence electrocyte shape, electrical properties, and potentially the morphology of penetrations and surface membrane invaginations.

To conclude, this study examined the expression correlates of a hyper-variable phenotype in a rapidly diversified genus of mormyrid electric fish. We examined DGE between taxa exhibiting variability along three major axes of variation that characterize EOD differences within *Paramormyrops* and among mormyrids: duration, polarity, and complexity. We found that gene expression in EOs among closely related species is largely similar, but patterns of DGE between EOs is primarily restricted to four broad functional sets: (1) cytoskeletal and sarcomeric proteins, (2) cation homeostasis, (3) lipid metabolism and (4) proteins that interact with the ECM. Our results suggest specific candidate genes that are likely

to influence the size, shape and architecture of electrocytes for future research on gene function and molecular pathways that underlie EOD variation in mormyrid electric fish.

APPENDIX

Table 1.1. Phenotypic and collection information of the samples studied.

Tag No.	OTU	Phenotypes per EOD feature			CUMV Accession Number	Site Name	Lat, Long	SL (mm)	Sex
		Duration	Complexity	Polarity (diameter of penetrations)					
PKINGN_6898	<i>P. kingsleyae</i> (N-type)	long EOD	biphasic	NA (no penetrations)	95184	Bikagala Creek	-2.20, 11.56	131	M
PKINGN_6900	<i>P. kingsleyae</i> (N-type)	long EOD	biphasic	NA (no penetrations)	95184	Bikagala Creek	-2.20, 11.56	145	M
PKINGN_6901	<i>P. kingsleyae</i> (N-type)	long EOD	biphasic	NA (no penetrations)	95184	Bikagala Creek	-2.20, 11.56	149	M
PKINGP_6716	<i>P. kingsleyae</i> (P-type)	long EOD	triphasic	small penetrations	95183	Mouvanga Creek	-2.33, 11.69	92.5	J/F
PKINGP_6718	<i>P. kingsleyae</i> (P-type)	long EOD	triphasic	small penetrations	95183	Mouvanga Creek	-2.33, 11.69	91	J/F
PMAG1_6780	<i>P. sp.</i> 'magnostipes type I'	long EOD	triphasic	large penetrations	95155	Mouvanga Creek	-2.33, 11.69	107	M
PMAG1_6787	<i>P. sp.</i> 'magnostipes type I'	long EOD	triphasic	large penetrations	95155	Mouvanga Creek	-2.33, 11.69	97.5	F
PMAG2_6768	<i>P. sp.</i> 'magnostipes type II'	long EOD	triphasic	small penetrations	95155	Mouvanga Creek	-2.33, 11.69	93	M
PMAG2_6769	<i>P. sp.</i> 'magnostipes type II'	long EOD	triphasic	small penetrations	95155	Mouvanga Creek	-2.33, 11.69	124	M

Table 1.1 (cont'd)

PSN3_6739	<i>P. sp.</i> 'SN3'	short EOD	biphasic	NA (no penetrations)	Uncatalogued	Mouvanga Creek	-2.33, 11.69	73	J/F
PSN3_6742	<i>P. sp.</i> 'SN3'	short EOD	biphasic	NA (no penetrations)	95173	Mouvanga Creek	-2.33, 11.69	70	J/F

CUMV = Cornell University Museum of Vertebrates, EOD = electric organ discharge, F= female, J = juvenile, Lat = Latitude, Long = Longitude, M = male, NA = not applicable, OTU = operational taxonomic unit, SL = standard length.

Table 1.2. All ten possible pairwise DGE comparisons with the total number of DEG and enriched GO terms for each. Also indicated is whether each comparison is informative for contrasting each EOD feature. The phenotypes for waveform polarity can only be contrasted in comparisons where both OTUs have penetrations. Informative comparisons for each EOD feature (Set A') are marked with an * in the column of the EOD feature they contrasted.

Comparison		Contrast			DEG genes	enriched GO terms		
OTU #1	OTU #2	Duration	Complexity	Polarity		BP	CC	MF
<i>P. kingsleyae</i> (N-type)	<i>P. kingsleyae</i> (P-type)	no	yes*	NA (no)	530	46	12	17
<i>P. kingsleyae</i> (N-type)	<i>P. sp.</i> 'magnostipes type I'	no	yes	NA (no)	1542	76	15	37
<i>P. kingsleyae</i> (N-type)	<i>P. sp.</i> 'magnostipes type II'	no	yes	NA (no)	1174	71	16	25
<i>P. kingsleyae</i> (N-type)	<i>P. sp.</i> 'SN3'	yes*	no	NA (no)	507	52	4	13
<i>P. kingsleyae</i> (P-type)	<i>P. sp.</i> 'magnostipes type I'	no	no	yes*	1322	40	12	25
<i>P. kingsleyae</i> (P-type)	<i>P. sp.</i> 'magnostipes type II'	no	no	no	719	47	10	24
<i>P. kingsleyae</i> (P-type)	<i>P. sp.</i> 'SN3'	yes	yes	NA (no)	385	33	3	14
<i>P. sp.</i> 'magnostipes type I'	<i>P. sp.</i> 'magnostipes type II'	no	no	yes	9	5	1	1
<i>P. sp.</i> 'magnostipes type I'	<i>P. sp.</i> 'SN3'	yes	yes	NA (no)	1053	43	6	27
<i>P. sp.</i> 'magnostipes type II'	<i>P. sp.</i> 'SN3'	yes	yes	NA (no)	489	40	9	16

BP = biological process, CC = cellular component, DEG = differentially expressed genes, DGE = differential gene expression, GO = gene ontology, NA = not applicable, MF = molecular function, OTU = operational taxonomic unit

Table 1.3. Total number of upregulated genes and enriched GO terms in Set C for each EOD feature, phenotype and ontology.

EOD feature	Phenotype	Upregulated genes	Protein-coding (%)	enriched GO terms		
				BP	CC	MF
Duration	short EODs	140	122 (87)	18	3	5
Duration	long EODs	43	34 (79)	10	1	2
Polarity	small penetrations	99	89 (90)	3	1	3
Polarity	large penetrations	55	40 (73)	6	1	3
Complexity	biphasic	110	81 (74)	7	1	2
Complexity	triphasic	35	25 (71)	7	0	3

BP = biological process, CC = cellular component, EOD = electric organ discharge, GO = gene ontology, MF = molecular function

Table 1.4. Selected DEG in Set C for waveform duration by “general” functional class and EOD phenotype, and highlights of their predicted function.

Pking_Entrez_geneID	Pking_gene_name	Pking_gene_symbol	“general” functional class	upregulated in phenotype	Highlights of Predicted Function (edited from UniProt)
111834716	<i>Kv channel-interacting protein 4</i>	LOC111834716	Cation homeostasis	short EODs	Regulatory subunit of Kv4/D (Shal)-type voltage-gated rapidly inactivating A-type potassium channels. Regulates channel density, inactivation kinetics and rate of recovery from inactivation in a calcium-dependent and isoform-specific manner
111836747	<i>potassium channel subfamily K member 5-like</i>	LOC111836747	Cation homeostasis	short EODs	pH-dependent, voltage insensitive, outwardly rectifying potassium channel. Outward rectification is lost at high external K ⁺ concentrations
111857989	<i>chloride intracellular channel protein 5-like</i>	LOC111857989	Cation homeostasis	short EODs	Can insert into membranes and form poorly selective ion channels that may also transport chloride ions. May play a role in the regulation of transepithelial ion absorption and secretion
111833088	<i>myosin-7-like</i>	LOC111833088	Cytoskeletal & sarcomeric	long EODs	Myosins are actin-based motor molecules with ATPase activity essential for muscle contraction
111856289	<i>pleckstrin homology-like domain family B member 1</i>	LOC111856289	Cytoskeletal & sarcomeric	long EODs	GO BP: regulation of microtubule cytoskeleton organization
111842483	<i>parvalbumin-2</i>	LOC111842483	Cytoskeletal & sarcomeric	short EODs	In muscle, parvalbumin is thought to be involved in relaxation after contraction. It binds two calcium ions
111846153	<i>troponin I, slow skeletal muscle-like</i>	LOC111846153	Cytoskeletal & sarcomeric	short EODs	Inhibitory subunit of troponin, the thin filament regulatory complex which confers calcium-sensitivity to striated muscle actomyosin ATPase activity
111851695	<i>keratin, type II cytoskeletal 8-like</i>	LOC111851695	Cytoskeletal & sarcomeric	short EODs	Together with KRT19, helps to link the contractile apparatus to dystrophin at the costameres of striated muscle
111856036	<i>parvalbumin-2-like</i>	LOC111856036	Cytoskeletal & sarcomeric	short EODs	In muscle, parvalbumin is thought to be involved in relaxation after contraction. It binds two calcium ions
111860236	<i>tropomyosin alpha-1 chain-like</i>	LOC111860236	Cytoskeletal & sarcomeric	short EODs	Binds to actin filaments in muscle and non-muscle cells. Plays a central role, in association with the troponin complex, in the calcium dependent regulation of vertebrate striated muscle contraction. In non-muscle cells is implicated in stabilizing cytoskeleton actin filaments.

Table 1.4 (cont'd)

111845490	<i>matrilin 2</i>	<i>matn2</i>	Extracellular matrix	long EODs	Involved in matrix assembly
111860169	<i>thrombospondin-4-like</i>	LOC111860169	Extracellular matrix	long EODs	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions and is involved in various processes including cellular proliferation, migration, adhesion and attachment
111860877	<i>collagenase 3-like</i>	LOC111860877	Extracellular matrix	short EODs	Plays a role in the degradation of extracellular matrix proteins
111834720	<i>protein EFR3 homolog B-like</i>	LOC111834720	Lipid metabolism	short EODs	Component of a complex required to localize phosphatidylinositol 4-kinase (PI4K) to the plasma membrane. The complex acts as a regulator of phosphatidylinositol 4-phosphate (PtdIns4P) synthesis
111840357	<i>PTB domain-containing engulfment adapter protein 1-like</i>	LOC111840357	Lipid metabolism	short EODs	Modulates cellular glycosphingolipid and cholesterol transport
111846286	<i>retinoic acid receptor responder protein 3-like</i>	LOC111846286	Lipid metabolism	short EODs	Catalyzes the calcium-independent hydrolysis of acyl groups in various phosphatidylcholines (PC) and phosphatidylethanolamine (PE)
111847640	<i>phosphatidylinositol 3-kinase regulatory subunit gamma-like</i>	LOC111847640	Lipid metabolism	short EODs	Binds to activated (phosphorylated) protein-tyrosine kinases through its SH2 domain and regulates their kinase activity
111852373	<i>proto-oncogene c-Fos-like</i>	LOC111852373	Lipid metabolism	short EODs	In growing cells, activates phospholipid synthesis, possibly by activating CDS1 and PI4K2A
111857713	<i>HRAS-like suppressor 3</i>	LOC111857713	Lipid metabolism	short EODs	Catalyzes the calcium-independent hydrolysis of acyl groups in various phosphatidylcholines (PC) and phosphatidylethanolamine (PE)
111860935	<i>fatty aldehyde dehydrogenase-like</i>	LOC111860935	Lipid metabolism	short EODs	Catalyzes the oxidation of medium and long chain aliphatic aldehydes to fatty acids

BP = biological process, DEG = differentially expressed genes, EOD = electric organ discharge, GO = gene ontology

Table 1.5. Selected DEG in Set C for waveform polarity, by “general” functional class and EOD phenotype, and highlights of their expected function.

Pking_Entrez_geneID	Pking_gene_name	Pking_gene_symbol	“general” functional class	upregulated in phenotype	Highlights of Predicted Function (edited from UniProt)
111840706	<i>G protein-activated inward rectifier potassium channel 1</i>	LOC111840706	Cation homeostasis	Small penetrations	This potassium channel is controlled by G proteins. Plays a crucial role in regulating the heartbeat
111853690	<i>protein kinase cGMP-dependent 1</i>	<i>prkg1</i>	Cation homeostasis	Small penetrations	Serine/threonine protein kinase. Numerous protein targets for PRKG1 phosphorylation are implicated in modulating cellular calcium. Proteins that are phosphorylated by PRKG1 regulate platelet activation and adhesion, smooth muscle contraction, cardiac function, gene expression
111843447	<i>myosin light chain 4-like</i>	LOC111843447	Cytoskeletal & sarcomeric	Large penetrations	Regulatory light chain of myosin. Does not bind calcium
111851664	<i>leucine rich repeat containing 10</i>	<i>lrrc10</i>	Cytoskeletal & sarcomeric	Large penetrations	May play important roles in cardiac development and/or cardiac function
111856907	<i>Wiskott-Aldrich syndrome protein family member 3-like</i>	LOC111856907	Cytoskeletal & sarcomeric	Large penetrations	Plays a role in the regulation of cell morphology and cytoskeletal organization. Required in the control of cell shape
111834243	<i>desmin-like</i>	LOC111834243	Cytoskeletal & sarcomeric	Small penetrations	Muscle-specific type III intermediate filament essential for proper muscular structure and function. Plays a crucial role in maintaining the structure of sarcomeres. May act as a sarcomeric microtubule-anchoring protein
111843225	<i>protein kinase C alpha type-like</i>	LOC111843225	Cytoskeletal & sarcomeric	Small penetrations	Regulates cardiomyocyte function by phosphorylating cardiac troponin T (TNNT2/CTNT), which induces significant reduction in actomyosin ATPase activity, myofilament calcium sensitivity and myocardial contractility
111848393	<i>PH and SEC7 domain-containing protein 1-like</i>	LOC111848393	Cytoskeletal & sarcomeric	Small penetrations	Induces cytoskeletal remodeling

Table 1.5 (cont'd)

111849608	<i>CDC42 effector protein 3</i>	<i>cdc42ep3</i>	Cytoskeletal & sarcomeric	Small penetrations	Probably involved in the organization of the actin cytoskeleton. May act downstream of CDC42 to induce actin filament assembly leading to cell shape changes
111853190	<i>protein phosphatase 1 regulatory subunit 12B-like</i>	LOC111853190	Cytoskeletal & sarcomeric	Small penetrations	Regulates myosin phosphatase activity. Augments Ca ²⁺ sensitivity of the contractile apparatus
111856340	<i>calponin-1-like</i>	LOC111856340	Cytoskeletal & sarcomeric	Small penetrations	Thin filament-associated protein that is implicated in the regulation and modulation of smooth muscle contraction. It is capable of binding to actin, calmodulin, troponin C and tropomyosin. The interaction of calponin with actin inhibits the actomyosin Mg-ATPase activity
111856797	<i>myosin light chain 3-like</i>	LOC111856797	Cytoskeletal & sarcomeric	Small penetrations	Regulatory light chain of myosin. Does not bind calcium
111838718	<i>inter-alpha-trypsin inhibitor heavy chain H5-like</i>	LOC111838718	Extracellular matrix	Small penetrations	inter-alpha-trypsin inhibitors usually interact with hyaluronan
111841241	<i>epiphycan-like</i>	LOC111841241	Extracellular matrix	Small penetrations	May have a role in bone formation and also in establishing the ordered structure of cartilage through matrix organization
111844627	<i>collagen alpha-1(V) chain-like</i>	LOC111844627	Extracellular matrix	Small penetrations	Type V collagen is a member of group I collagen (fibrillar forming collagen). It is a minor connective tissue component of nearly ubiquitous distribution. Type V collagen binds to DNA, heparan sulfate, thrombospondin, heparin, and insulin
111853425	<i>thrombospondin 2</i>	<i>thbs2</i>	Extracellular matrix	Small penetrations	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions
111854264	<i>collagen type IX alpha 2 chain</i>	<i>col9a2</i>	Extracellular matrix	Small penetrations	Structural component of hyaline cartilage and vitreous of the eye
111857302	<i>anthrax toxin receptor 1-like</i>	LOC111857302	Extracellular matrix	Small penetrations	Interacts with extracellular matrix proteins and with the actin cytoskeleton. Mediates adhesion of cells to type 1 collagen and gelatin, reorganization of the actin cytoskeleton and promotes cell spreading
111857834	<i>collagen alpha-4(VI) chain-like</i>	LOC111857834	Extracellular matrix	Small penetrations	Collagen VI acts as a cell-binding protein

Table 1.5 (cont'd)

111859912	<i>cell migration-inducing and hyaluronan-binding protein-like</i>	LOC111859912	Extracellular matrix	Small penetrations	Mediates depolymerization of hyaluronic acid (HA) via the cell membrane-associated clathrin-coated pit endocytic pathway. Binds to hyaluronic acid
111834720	<i>protein EFR3 homolog B-like</i>	LOC111834720	Lipid metabolism	Large penetrations	Component of a complex required to localize phosphatidylinositol 4-kinase (PI4K) to the plasma membrane. The complex acts as a regulator of phosphatidylinositol 4-phosphate (PtdIns4P) synthesis
111840084	<i>cytosolic phospholipase A2-like</i>	LOC111840084	Lipid metabolism	Small penetrations	Selectively hydrolyzes arachidonyl phospholipids in the sn-2 position releasing arachidonic acid
111853114	<i>alkaline ceramidase 2-like</i>	LOC111853114	Lipid metabolism	Small penetrations	Hydrolyzes the sphingolipid ceramide into sphingosine and free fatty acid

DEG = differentially expressed genes, EOD = electric organ discharge.

Table 1.6. Selected DEG in Set C for waveform complexity, by “general” functional class and EOD phenotype, and highlights of their expected function.

Pking_Entrez_geneID	Pking_gene_name	Pking_gene_symbol	“general” functional class	upregulated in phenotype	Highlights of Predicted Function (edited from UniProt)
111838181	<i>solute carrier family 9 member A7</i>	<i>slc9a7</i>	Cation homeostasis	Biphasic	Protein: Sodium/hydrogen exchanger 7. Gene: SLC9A7. Mediates electroneutral exchange of protons for Na ⁺ and K ⁺ across endomembranes
111848312	<i>voltage-dependent calcium channel gamma-1 subunit-like</i>	LOC111848312	Cation homeostasis	Triphasic	Regulatory subunit of the voltage-gated calcium channel that gives rise to L-type calcium currents in skeletal muscle. Regulates channel inactivation kinetics
111841270	<i>family with sequence similarity 110 member C</i>	<i>fam110c</i>	Cytoskeletal & sarcomeric	Biphasic	May play a role in microtubule organization
111845832	<i>5'-AMP-activated protein kinase subunit gamma-2-like</i>	LOC111845832	Cytoskeletal & sarcomeric	Biphasic	AMP/ATP-binding subunit of AMP-activated protein kinase (AMPK). Acts as a regulator of cellular polarity by remodeling the actin cytoskeleton; probably by indirectly activating myosin
111850616	<i>transmembrane protein 47-like</i>	LOC111850616	Cytoskeletal & sarcomeric	Biphasic	Regulates cell junction organization in epithelial cells. May regulate F-actin polymerization
111851223	<i>FYVE, RhoGEF and PH domain-containing protein 4-like</i>	LOC111851223	Cytoskeletal & sarcomeric	Biphasic	Plays a role in regulating the actin cytoskeleton and cell shape
111857398	<i>protein-methionine sulfoxide oxidase mical2b-like</i>	LOC111857398	Cytoskeletal & sarcomeric	Biphasic	Promotes depolymerization of F-actin
111857697	<i>leiomodlin 2</i>	<i>lmod2</i>	Cytoskeletal & sarcomeric	Biphasic	Mediates nucleation of actin filaments and thereby promotes actin polymerization. Plays a role in the regulation of actin filament length. Required for normal sarcomere organization in the heart, and for normal heart function

Table 1.6 (cont'd)

111854588	<i>capping protein regulator and myosin 1 linker 3</i>	<i>carmil3</i>	Cytoskeletal & sarcomeric	Triphasic	No info for CARMIL3, but CARMIL2 is a cell membrane-cytoskeleton-associated protein that plays a role in the regulation of actin polymerization at the barbed end of actin filaments. Enhances actin polymerization
111841398	<i>inter-alpha-trypsin inhibitor heavy chain 3</i>	<i>itih3</i>	Extracellular matrix	Biphasic	May act as a carrier of hyaluronan in serum or as a binding protein between hyaluronan and other matrix proteins
111841399	<i>inter-alpha-trypsin inhibitor heavy chain H3-like</i>	LOC111841399	Extracellular matrix	Biphasic	May act as a carrier of hyaluronan in serum or as a binding protein between hyaluronan and other matrix proteins
111853010	<i>ependymin-like</i>	LOC111853010	Extracellular matrix	Triphasic	GO MF: calcium ion binding. GO BP: cell-matrix adhesion
111853814	<i>epiphycan-like</i>	LOC111853814	Extracellular matrix	Triphasic	May have a role in bone formation and also in establishing the ordered structure of cartilage through matrix organization

BP = biological process, DEG = differentially expressed genes, EOD = electric organ discharge, GO = gene ontology, MF = molecular function.

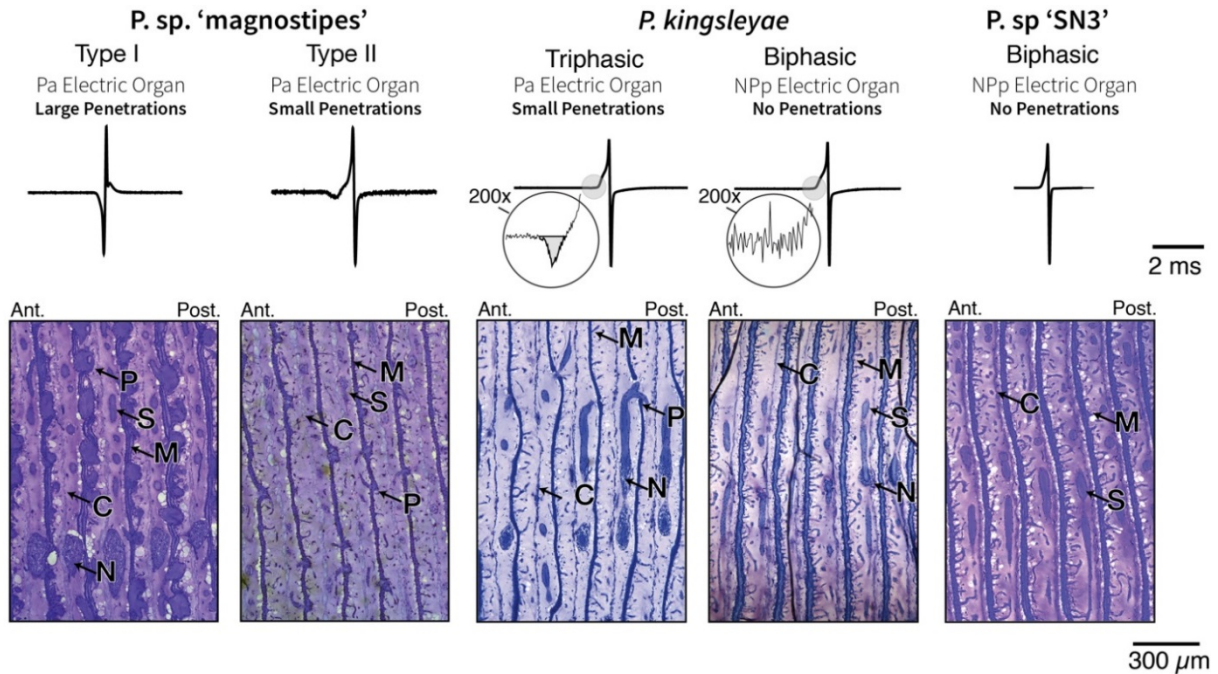


Figure 1.1. Electric organ discharge (EOD) diversity and electric organ anatomy in *Paramormyrops*. EOD traces from specimens in this study and representative parasagittal sections of the five *Paramormyrops* operational taxonomic units (OTUs) considered in this study. 200x magnification on *P. kingsleyae* EODs reveals a P0 phase on triphasic EODs only. Individuals with triphasic EODs all have penetrations, whereas individuals with biphasic EODs do not. OTUs with 'inverted' polarity triphasic EODs have large penetrations compared to OTUs with normal polarity triphasic EODs. Ant. = anterior, C = connective tissue septa, N = nerve, NPp = non-penetrating, posteriorly innervated stalks, M = microstalklets (profusely branched stalks), P = penetrations, Pa = penetrating, anteriorly innervated stalks, Post. = posterior, S = stalks. Image from *P. kingsleyae* biphasic originally appeared in Gallant et al. 2011.

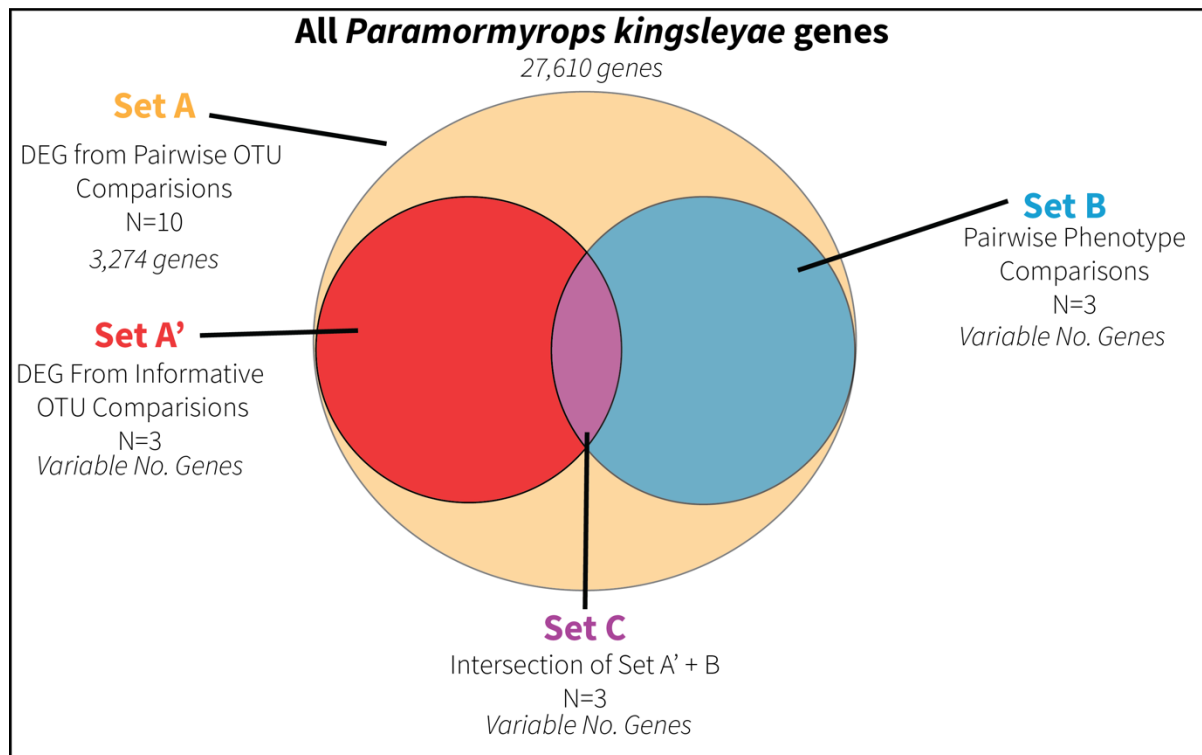


Figure 1.2. Diagram of how we constructed the lists of upregulated genes of Set C. DEG = differentially expressed genes, N = number of comparisons made for each set, OTU = operational taxonomic unit

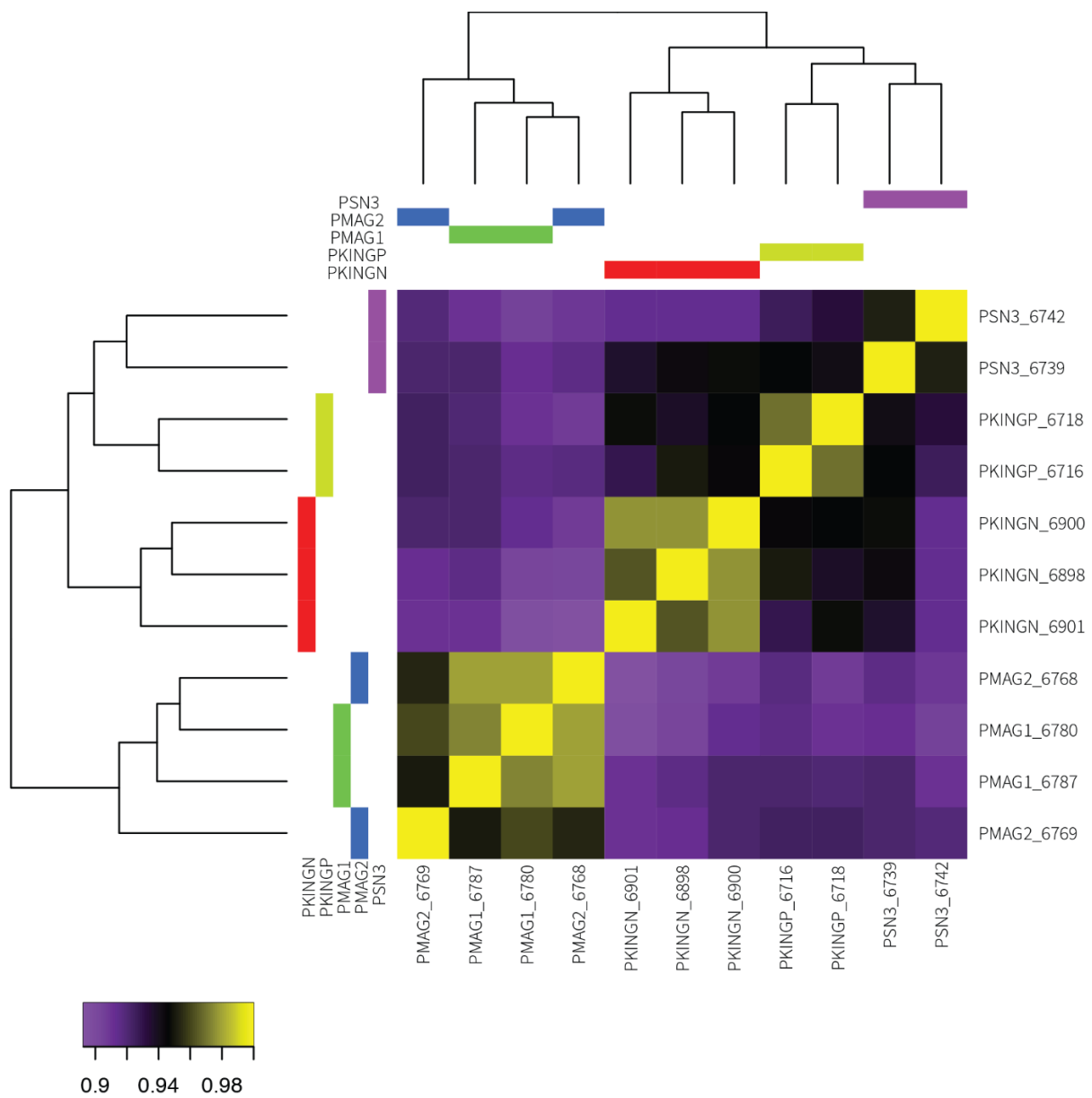


Figure 1.3. Heatmap of sample by sample correlations in gene expression, and the inferred relationships among OTUs from these expression correlation values. OTU = operational taxonomic unit.

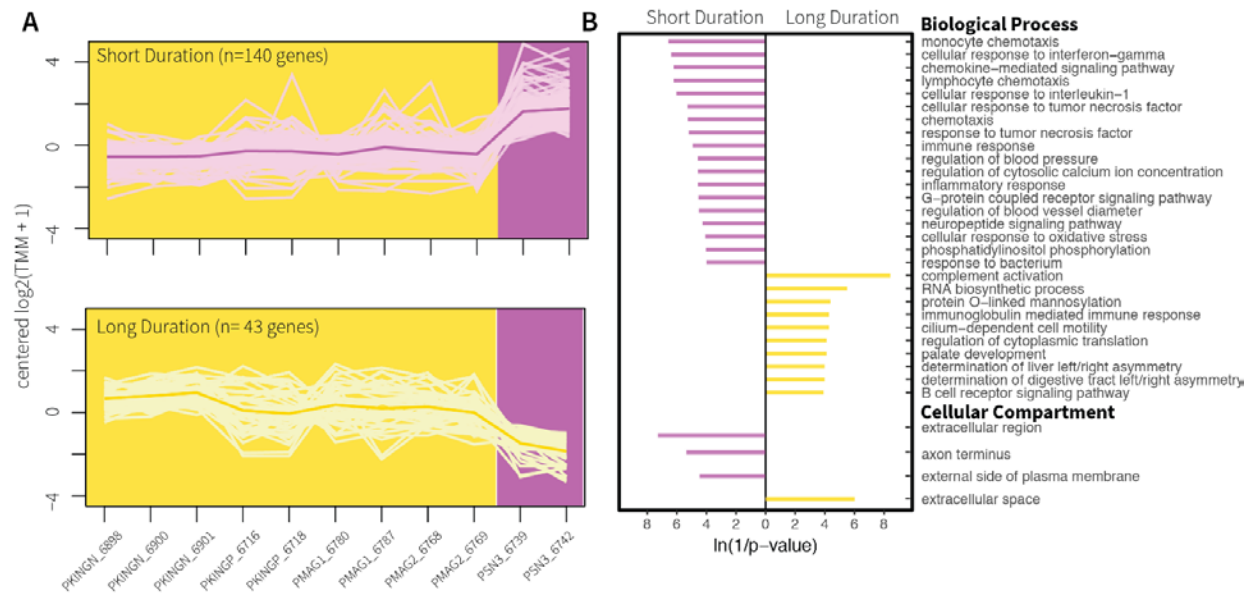


Figure 1.4. Set C for waveform duration. A) Expression patterns of Set C genes for the waveform duration phenotypes short EODs (purple background) and long EODs (yellow background). Samples are sorted alphabetically on the X axis. The lines connect transformed gene expression values across all samples; light-color lines represent one gene, the dark-color line is the average expression pattern of all genes. B) Gene Ontology (GO) terms for Biological Process and Cellular Component found enriched in the gene lists from (A). The X axis shows transformed p-values, the longer a bar the smaller its p-value. The direction and color of a bar indicate the phenotype in which the GO term is enriched [same color code as (A)].

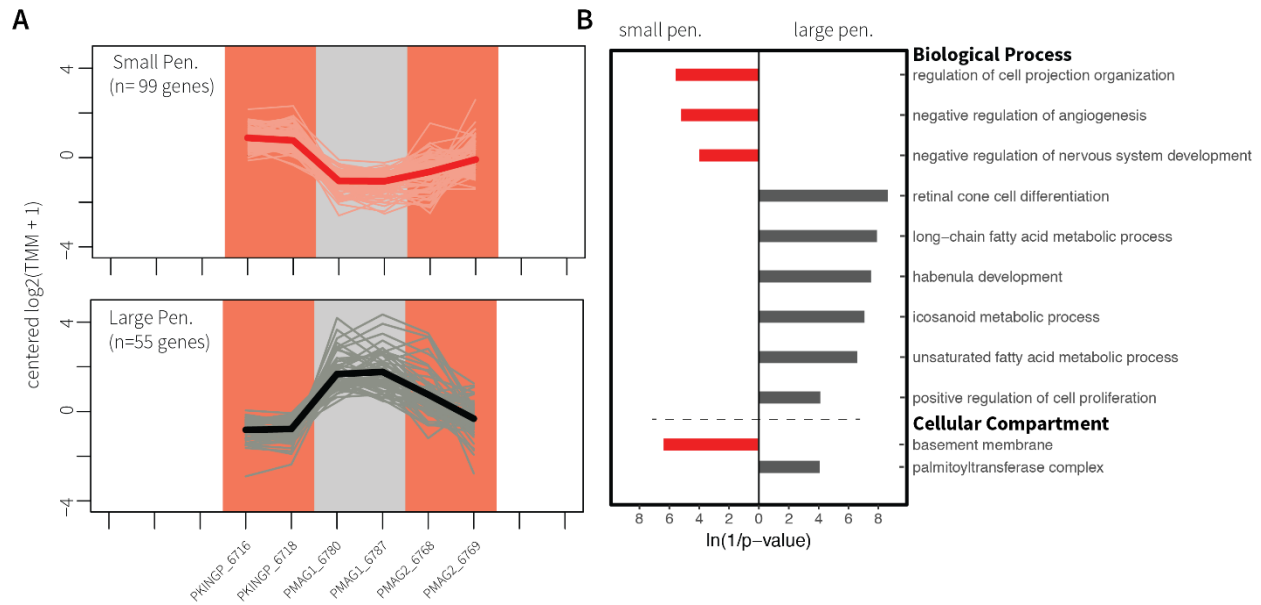


Figure 1.5. Set C for waveform polarity. A) Expression patterns of Set C genes for the waveform polarity phenotypes small penetrations (red background) and large penetrations (grey background). Samples are sorted alphabetically on the X axis. The lines connect transformed gene expression values across all samples; light-color lines represent one gene, the dark-color line is the average expression pattern of all genes. B) Gene Ontology (GO) terms for Biological Process and Cellular Component found enriched in the gene lists from (A). The X axis shows transformed p-values, the longer a bar the smaller its p-value. The direction and color of a bar indicate the phenotype in which the GO term is enriched [same color code as (A)]. Pen = penetrations.

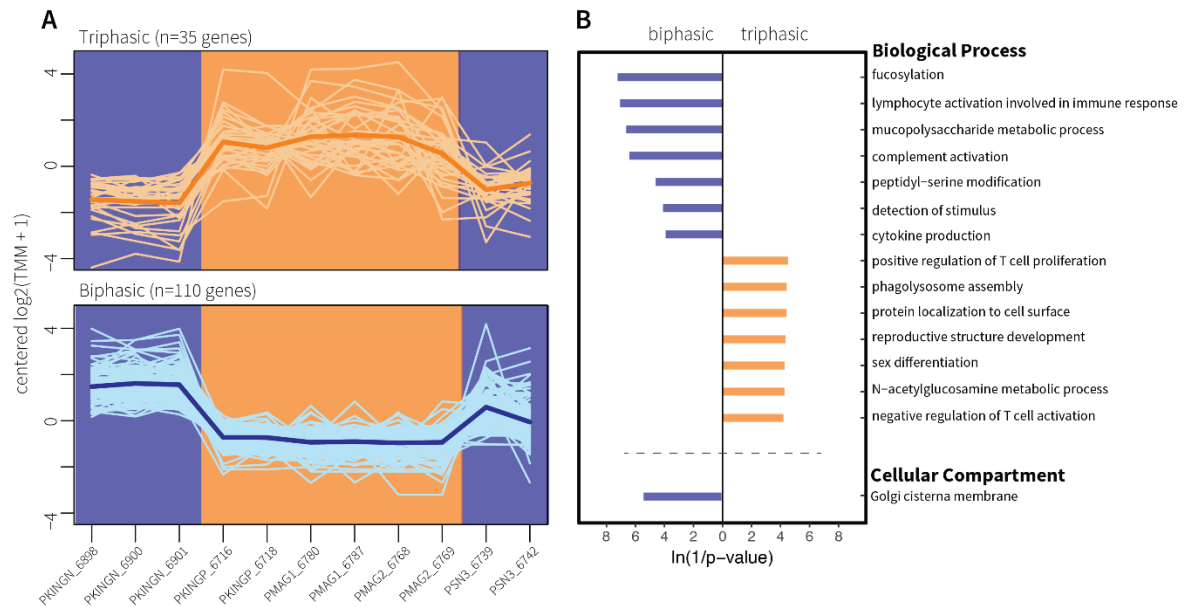


Figure 1.6. Set C for waveform complexity. A) Expression patterns of Set C genes for the waveform complexity phenotypes triphasic (orange background) and biphasic (blue background). Samples are sorted alphabetically on the X axis. The lines connect transformed gene expression values across all samples; light-color lines represent one gene, the dark-color line is the average expression pattern of all genes. B) Gene Ontology (GO) terms for Biological Process and Cellular Component found enriched in the gene lists from (A). The X axis shows transformed p-values, the longer a bar the smaller its p-value. The direction and color of a bar indicate the phenotype in which the GO term is enriched [same color code as (A)].

REFERENCES

REFERENCES

- Abolins-Abols, M., E. Kornobis, P. Ribeca, K. Wakamatsu, M. P. Peterson, E. D. Ketterson, and B. Milá. 2018. Differential gene regulation underlies variation in melanic plumage coloration in the dark-eyed junco (*Junco hyemalis*). *Mol. Ecol.* 27:4501–4515.
- Alexa, A., J. Rahnenführer, and T. Lengauer. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22:1600–1607. Oxford University Press.
- Alvarez, M., A. W. Schrey, and C. L. Richards. 2015. Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Mol. Ecol.* 24:710–725.
- Alves-Gomes, J., and C. D. Hopkins. 1997. Molecular insights into the phylogeny of mormyrid form fishes and the evolution of their electric organs. *Brain Behav. Evol.* 49:324–350.
- Arnegard, M. E., S. M. Bogdanowicz, and C. D. Hopkins. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. *Evolution* 59:324–343.
- Arnegard, M. E., B. S. Jackson, and C. D. Hopkins. 2006. Time-domain signal divergence and discrimination without receptor modification in sympatric morphs of electric fishes. *J. Exp. Biol.* 209:2182–2198.
- Arnegard, M. E., P. B. McIntyre, L. J. Harmon, M. L. Zelditch, W. G. R. Crampton, J. K. Davis, J. P. Sullivan, S. Lavoué, and C. D. Hopkins. 2010a. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *Am. Nat.* 176:335–356.
- Arnegard, M. E., D. J. Zwickl, Y. Lu, and H. H. Zakon. 2010b. Old gene duplication facilitates origin and diversification of an innovative communication system--twice. *Proc. Natl. Acad. Sci. U.S.A.* 107:22172–22177.
- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29.
- Bartels, E. 1971. Depolarization of electroplax membrane in calcium-free ringer's solution. *J. Membr. Biol.* 5:121–132.
- Bass, A. H. 1986. Species differences in electric organs of mormyrids: Substrates for species-typical electric organ discharge waveforms. *J. Comp. Neurol.* 244:313–330.
- Bass, A. H., J.-P. Denizot, and M. A. Marchaterre. 1986. Ultrastructural features and hormone-dependent sex differences of mormyrid electric organs. *J. Comp. Neurol.* 254:511–528.
- Bass, A. H., and C. D. Hopkins. 1985. Hormonal control of sex differences in the electric organ discharge (EOD) of mormyrid fishes. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 156:587–604.

- Bass, A. H., and C. D. Hopkins. 1983. Hormonal control of sexual differentiation: changes in electric organ discharge waveform. *Science* 220:971–974.
- Bass, A. H., and S. F. Volman. 1987. From behavior to membranes: testosterone-induced changes in action potential duration in electric organs. *Proc. Natl. Acad. Sci. U.S.A.* 84:9295–9298.
- Bennett, M. V. L. 1971. Electric Organs. Pp. 347–491 *in* W. S. Hoar and D. J. Randall, eds. *Fish Physiology*. Academic Press, London.
- Bennett, M. V. L., and H. Grundfest. 1961. Studies on the morphology and electrophysiology of electric organs. III. Electrophysiology of electric organs in mormyrids. Pp. 113–35 *in* C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, Amsterdam.
- Binns, D., E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan, and R. Apweiler. 2009. QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* 25:3045–3046. Oxford University Press.
- Blankers, T., K. P. Oh, A. Bombarely, and K. L. Shaw. 2018. The Genomic Architecture of a Rapid Island Radiation: Recombination Rate Variation, Chromosome Structure, and Genome Assembly of the Hawaiian Cricket *Laupala*. *Genetics* 209:1329–1344. *Genetics*.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Brawand, D., M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, F. W. Albert, U. Zeller, P. Khaitovich, F. Grützner, S. Bergmann, R. Nielsen, S. Pääbo, and H. Kaessmann. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–8.
- Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, R. F. Castillo Cajas, W. Diao, M. E. Maan, S. Paolucci, F. J. Weissing, L. van de Zande, A. Hoikkala, E. Geuverink, J. Jennings, M. Kankare, K. E. Knott, V. I. Tyukmaeva, C. Zoumadakis, M. G. Ritchie, D. Barker, E. Immonen, M. Kirkpatrick, M. Noor, M. C. Garcia, T. Schmitt, and M. Schilthuizen. 2012. What do we need to know about speciation? *Trends Ecol. Evol.* 27:27–39.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. BioMed Central.
- Caputi, A. A., B. A. Carlson, and O. Macadar. 2005. Electric Organs and Their Control. Pp. 410–451 *in* T. H. Bullock, Hopkins C.D., Popper A.N., and Fay R.R., eds. *Electroreception*. Springer Handbook of Auditory Research, vol 21, New York.
- Carlson, B. A. 2002. Electric signaling behavior and the mechanisms of electric organ discharge production in mormyrid fish. *J. Physiol. Paris* 96:405–419. Elsevier.
- Carlson, B. A., and J. R. Gallant. 2013. From sequence to spike to spark: evo-devo-neuroethology of electric communication in mormyrid fishes. *J. Neurogenet.* 27:106–129.
- Carlson, B. A., S. M. Hasan, M. Hollmann, D. B. Miller, L. J. Harmon, and M. E. Arnegard. 2011. Brain evolution triggers increased diversification of electric fishes. *Science* 332:583–586.

- Denizot, J. P., F. Kirschbaum, G. W. M. Westby, and S. Tsuji. 1982. On the development of the adult electric organ in the mormyrid fish *Pollimyrus isidori* (with special focus on the innervation). *J. Neurocytol.* 11:913–934.
- Ding, B., D. W. Daugherty, M. Husemann, M. Chen, A. E. Howe, and P. D. Danley. 2014. Quantitative genetic analyses of male color pattern and female mate choice in a pair of cichlid fishes of Lake Malawi, East Africa. *PLOS One* 9:1–22.
- Ferrari, M. B., and H. H. Zakon. 1993. Conductances contributing to the action-potential of *Sternopygus electrocytes*. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 173:281–292. Springer.
- Ferreira, P. G., S. Patalano, R. Chauhan, R. Ffrench-Constant, T. Gabaldón, R. Guigó, and S. Sumner. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14:R20.
- Fraser, J. R. E., T. C. Laurent, and U. B. G. Laurent. 1997. Hyaluronan: its nature, distribution, functions and turnover. *J. Intern. Med.* 242:27–33. John Wiley & Sons, Ltd (10.1111).
- Freedman, E. G., J. Olyarchuk, M. A. Marchaterre, and A. H. Bass. 1989. A temporal analysis of testosterone-induced changes in electric organs and electric organ discharges of mormyrid fishes. *J. Neurobiol.* 20:619–634.
- Gallant, J. R., M. E. Arnegard, J. P. Sullivan, B. A. Carlson, and C. D. Hopkins. 2011. Signal variation and its morphological correlates in *Paramormyrops kingsleyae* provide insight into the evolution of electrogenic signal diversity in mormyrid electric fish. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 197:799–817.
- Gallant, J. R., C. D. Hopkins, and D. L. Deitcher. 2012. Differential expression of genes and proteins between electric organ and skeletal muscle in the mormyrid electric fish *Brienomyrus brachyistius*. *J. Exp. Biol.* 215:2479–2494.
- Gallant, J. R., M. Losilla, C. Tomlinson, and W. C. Warren. 2017. The Genome and Adult Somatic Transcriptome of the Mormyrid Electric Fish *Paramormyrops kingsleyae*. *Genome Biol. Evol.* 9:3525–3530. Oxford University Press.
- Gallant, J. R., L. L. Traeger, J. D. Volkening, H. Moffett, P.-H. Chen, C. D. Novina, G. N. Phillips, R. Anand, G. B. Wells, M. Pinch, R. Guth, G. A. Unguez, J. S. Albert, H. H. Zakon, M. P. Samanta, and M. R. Sussman. 2014. Genomic basis for the convergent evolution of electric organs. *Science* 344:1522–1525.
- Ganss, B., and W. Hoffmann. 2009. Calcium-induced conformational transition of trout ependymins monitored by tryptophan fluorescence. *Open Biochem. J.* 3:14–17.
- Giorello, F. M., M. Feijoo, G. D’Elía, D. E. Naya, L. Valdez, J. C. Opazo, and E. P. Lessa. 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). *Mol. Ecol.* 27:3274–3286.
- Givnish, T. J. 2015. Adaptive radiation versus “radiation” and “explosive diversification”: Why conceptual distinctions are fundamental to understanding evolution. *New Phytol.* 207:297–303.

- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–52. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
- Grant, R. B., and P. R. Grant. 2003. What Darwin’s Finches Can Teach Us about the Evolutionary Origin and Regulation of Biodiversity. *Bioscience* 53:965–975.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–512. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
- Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome Res.* 27:1004–1015.
- Härer, A., A. Meyer, and J. Torres-Dowdall. 2018. Convergent phenotypic evolution of the visual system via different molecular routes: How Neotropical cichlid fishes adapt to novel light environments. *Evol. Lett.* 2:341–354.
- Hopkins, C. D. 1986. Behavior of Mormyridae. Pp. 527–576 in T. H. Bullock and W. Heiligenberg, eds. *Electroreception*. Wiley, New York.
- Hopkins, C. D. 1999a. Design features for electric communication. *J. Exp. Biol.* 202:1217–1228. *Co Biol.*
- Hopkins, C. D. 1981. On the diversity of electric signals in a community of mormyrid electric fish in West Africa. *Am. Zool.* 21:211–222. Oxford University Press.
- Hopkins, C. D. 1999b. Signal evolution in electric communication. Pp. 461– 491 in M. Hauser and M. Konishi, eds. *The design of animal communication*. MIT Press, Cambridge, MA.
- Hulsey, C. D. 2009. Cichlid genomics and phenotypic diversity in a comparative context. *Integr. Comp. Biol.* 49:618–629.
- Jeukens, J., S. Renaut, J. St-Cyr, A. W. Nolte, and L. Bernatchez. 2010. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol. Ecol.* 19:5389–5403.
- King, M. C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
- Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nat. Rev. Genet.* 5:288–298.

Kramer, B. 1974. Electric organ discharge interaction during interspecific agonistic behaviour in freely swimming mormyrid fish. *J. Comp. Physiol.* 93:203–235. Springer-Verlag.

Lamanna, F., F. Kirschbaum, and R. Tiedemann. 2014. De novo assembly and characterization of the skeletal muscle and electric organ transcriptomes of the African weakly electric fish *Campylomormyrus compressirostris* (Mormyridae, Teleostei). *Mol. Ecol. Resour.* 14:1222–1230.

Lamanna, F., F. Kirschbaum, I. Waurick, C. Dieterich, and R. Tiedemann. 2015. Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). *BMC Genomics* 16:668. BMC Genomics.

Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

Lamichhaney, S., F. Han, J. Berglund, C. Wang, M. S. Almén, M. T. Webster, B. R. Grant, P. R. Grant, and L. Andersson. 2016. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352:470–474. American Association for the Advancement of Science.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359. Nature Publishing Group.

Lavoué, S., M. E. Arnegard, J. P. Sullivan, and C. D. Hopkins. 2008. Petrocephalus of Odzala offer insights into evolutionary patterns of signal diversification in the Mormyridae, a family of weakly electrogenic fishes from Africa. *J. Physiol. Paris* 102:322–339. Elsevier.

Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. BioMed Central.

Lissmann, H. W., and K. E. Machin. 1958. The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J. Exp. Biol.* 35:451–486.

MacManes, M. D. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* 5:1–7.

Magalhaes, I. S., and O. Seehausen. 2010. Genetics of male nuptial colour divergence between sympatric sister species of a Lake Victoria cichlid fish. *J. Evol. Biol.* 23:914–924. John Wiley & Sons, Ltd (10.1111).

Möhres, F. P. 1957. Elektrische Entladungen im Dienste der Revierabgrenzung bei Fischen. *Naturwissenschaften* 44:431–432. Springer-Verlag.

Mullen, S. P., T. C. Mendelson, C. Schal, and K. L. Shaw. 2007. Rapid evolution of cuticular hydrocarbons in a species radiation of acoustically diverse Hawaiian crickets (Gryllidae: Trigonidiinae: Laupala). *Evolution* 61:223–231.

Nagel, R., F. Kirschbaum, and R. Tiedemann. 2017. Electric organ discharge diversification in mormyrid weakly electric fish is associated with differential expression of voltage-gated ion channel genes. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 203:183–195. Springer Berlin Heidelberg.

- Nigenda-Morales, S. F., Y. Hu, J. C. Beasley, H. A. Ruiz-Piña, D. Valenzuela-Galván, and R. K. Wayne. 2018. Transcriptomic analysis of skin pigmentation variation in the Virginia opossum (*Didelphis virginiana*). *Mol. Ecol.* 27:2680–2697.
- O’Quin, C. T., A. C. Drilea, R. B. Roberts, and T. D. Kocher. 2012. A Small Number of Genes Underlie Male Pigmentation Traits in Lake Malawi Cichlid Fishes. *J. Exp. Zool. Part B Mol. Dev. Evol.* 318:199–208.
- Paul, C., F. Kirschbaum, V. Mamonekene, and R. Tiedemann. 2016. Evidence for Non-neutral Evolution in a Sodium Channel Gene in African Weakly Electric Fish (*Campylomormyrus*, Mormyridae). *J. Mol. Evol.* 83:61–77. Springer New York LLC.
- Picq, S., J. Sperling, C. J. Cheng, B. A. Carlson, and J. R. Gallant. 2020. Genetic drift does not sufficiently explain patterns of electric signal variation among populations of the mormyrid electric fish *Paramormyrops kingsleyae*. *Evolution* 1–25.
- Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4:1–8.
- Robinson, M. D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.
- Schwartz, I. R., G. D. Pappas, and M. V. L. Bennett. 1975. The fine structure of electrocytes in weakly electric teleosts. *J. Neurocytol.* 4:87–114.
- Seehausen, O. 2006. African cichlid fish: A model system in adaptive radiation research. *Proc. Royal Soc. B* 273:1987–1998.
- Shaw, K. L. 2000. Further acoustic diversity in Hawaiian forests: two new species of Hawaiian cricket (Orthoptera: Gryllidae: Trigonidiinae: Laupala). *Zool. J. Linn. Soc.* 129:73–91.
- Shaw, K. L., and S. C. Lesnick. 2009. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl. Acad. Sci. U.S.A.* 106:9737–9742. National Academy of Sciences.
- Suárez-Castillo, E. C., and J. E. García-Arrarás. 2007. Molecular evolution of the ependymin protein family: a necessary update. *BMC Evol. Biol.* 7:23.
- Sullivan, J. P., S. Lavoué, M. E. Arnegard, and C. D. Hopkins. 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* 58:825–41.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2002. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution* 56:597–616. Society for the Study of Evolution.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2000. Molecular systematics of the African electric fishes (Mormyroidea: teleostei) and a model for the evolution of their electric organs. *J. Exp. Biol.* 203:665–683.

- Swapna, I., A. Ghezzi, J. M. York, M. R. Markham, D. B. Halling, Y. Lu, J. R. Gallant, and H. H. Zakon. 2018. Electrostatic Tuning of a Potassium Channel in Electric Fish. *Curr. Biol.* 28:2094-2102.e5. Cell Press.
- Szabo, T. 1960. Development of the Electric Organ of Mormyridae. *Nature* 188:760–762. Nature Publishing Group.
- Szabo, T. 1961. Les Organes Electriques des Mormyrides. Pp. 20–24 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, New York.
- The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47:D330–D338.
- The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47:D506–D515.
- von der Emde, G., M. Amey, J. Engelmann, S. Fetz, C. Folde, M. Hollmann, M. Metzen, and R. Pusch. 2008. Active electrolocation in *Gnathonemus petersii*: Behaviour, sensory performance, and receptor systems. *J. Physiol. Paris* 102:279–290. Elsevier Ltd.
- Wiley, C., C. K. Ellison, and K. L. Shaw. 2012. Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. *Proc. Royal Soc. B* 279:1203–1209. The Royal Society.
- Wilwert, J. L., M. Madhoun, Nisreen, and D. J. Coughlin. 2006. Parvalbumin correlates with relaxation rate in the swimming muscle of sheephead and kingfish. *J. Exp. Biol.* 209:227–237.
- Wu, C., I. MacLeod, and A. I. Su. 2013. BioGPS and MyGene.info: Organizing online, gene-centric information. *Nucleic Acids Res.* 41:D561–D565. Oxford University Press.
- Xin, J., A. Mark, C. Afrasiabi, G. Tsueng, M. Juchler, N. Gopal, G. S. Stupp, T. E. Putman, B. J. Ainscough, O. L. Griffith, A. Torkamani, P. L. Whetzel, C. J. Mungall, S. D. Mooney, A. I. Su, and C. Wu. 2016. High-performance web services for querying gene and variant annotation. *Genome Biol.* 17:91. BioMed Central.
- Young, R. L., M. H. Ferkin, N. F. Ockendon-Powell, V. N. Orr, S. M. Phelps, Á. Pogány, C. L. Richards-Zawacki, K. Summers, T. Székely, B. C. Trainor, A. O. Urrutia, G. Zachar, L. A. O’Connell, and H. A. Hofmann. 2019. Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* 116:201813775.
- Zakon, H. H., Y. Lu, D. J. Zwickl, and D. M. Hillis. 2006. Sodium channel genes and the evolution of diversity in communication signals of electric fishes: Convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 103:3675–3680.

CHAPTER 2: MECHANISTIC INSIGHTS INTO GENE EXPRESSION CHANGES AND ELECTRIC ORGAN DISCHARGE ELONGATION IN MORMYRID ELECTRIC FISH

Introduction

Despite little ecological differentiation, African weakly electric fish (Mormyridae) hold one of the highest rates of speciation among ray-finned fishes (Carlson et al. 2011; Rabosky et al. 2013). It is suspected that a crucial factor of their extraordinary diversification is divergence in their electric organ discharges (EODs), which have evolved faster than size, morphology and trophic ecology (Arnegard et al. 2010a). These electric signals, crucial to electrolocation (Lissmann and Machin 1958; von der Emde et al. 2008) and communication (Möhres 1957; Kramer 1974), are effective mechanisms of prezygotic isolation (Hopkins and Bass 1981; Arnegard et al. 2006), and often are the most reliable method for identifying species (Hopkins 1981; Bass 1986; Arnegard et al. 2005).

Species-specific mormyrid EOD waveforms vary primarily in their complexity (number of phases), polarity (alterations to waveform shape caused by large a P0), and duration (Hopkins 1999b; Sullivan et al. 2002), the latter being the most variable factor between species (Arnegard and Hopkins 2003; Gallant et al. 2011) and the basis for recognition of species-specific signals (Hopkins and Bass 1981; Arnegard et al. 2006). EOD duration differs up to 100x among species (Hopkins 1999a), ranging from 0.2 to > 15 ms (Lavoué et al. 2008); it is often dimorphic (Bass and Hopkins 1983; Hopkins 1999a), with differences between male and female adult EODs often exacerbated during the wet-season breeding period (Hopkins and Bass 1981).

The distinctive features of the EOD waveform are a consequence of the physiological, morphological, and ultrastructural characteristics of the electrocytes (Carlson 2002). Mormyrid EODs are produced by the synchronized discharge of the electrocytes, the constitutive cells of the electric organ (EO), located in the caudal peduncle (Hopkins 1986). The EOD from an individual electrocyte is composed of the sum of the action potentials (APs) from different parts of its excitable membrane. The two largest phases of a typical mormyrid EOD, called P1 and P2, are generated by the combined APs of

the posterior and anterior faces of the electrocyte membrane (Bennett and Grundfest 1961). These two membranes are thought to receive the depolarization signal simultaneously, but ions begin flowing through the posterior face first (Stoddard and Markham 2008). The reasons for this delay could be physiological (ion channels) and/or due to greater capacitance of the more intricately folded anterior face; however, the majority of the two APs still overlap (Fig. 2.1). Therefore, the resulting EOD is the net current of overlapping in- and outgoing ionic currents at the two faces of the membrane (reviewed in (Markham 2013; Dunlap et al. 2017)). This model is thought to apply broadly to biphasic EODs from mormyrids and the Gymnotiformes (a diverse clade of independently evolved neotropical weakly electric fish). Biphasic EODs have only P1 and P2, but several mormyrids produce triphasic waveforms with an additional, initial phase, P0. Their electrocytes have stalks that penetrate the electrocyte (Pa configuration), P0 is produced when current flows through these penetrating stalks (more details in (Bennett and Grundfest 1961; Szabo 1961; Bennett 1971)).

EOD duration is thought to be regulated by both morphological and physiological aspects. Electrocytes with a larger membrane surface area have been correlated with EODs of longer duration (Bennett 1971; Paul et al. 2015), especially when the area increase is more prominent in the anterior face of the membrane (Bass et al. 1986; Freedman et al. 1989; Paul et al. 2015; Nguyen et al. 2020). Presumably, the larger area raises the capacitance of the membrane, and thus increases the time it takes to start depolarizing (Bass et al. 1986; Bass and Volman 1987; Stoddard and Markham 2008). If this delay is more pronounced in the anterior face of the membrane, the net effect should be to extend EOD duration (Fig. 2.1). In addition, changes to the abundance or kinetics of the ion channels in the electrocyte membrane are expected to also modulate EOD duration. In gymnotiforms, it has been demonstrated that the main electrocyte ionic currents are generated by voltage-gated Na^+ and K^+ currents (e.g. (Shenkel and Sigworth 1991; Ferrari and Zakon 1993)) and it is broadly expected that the same holds true in mormyrid electrocytes, where both Na^+ (Zakon et al. 2006; Arnegard et al. 2010b)

and K⁺ (Nagel et al. 2017; Swapna et al. 2018) voltage-gated channels are expressed, and differential expression of ion channel genes has been linked with interspecific differences in EOD waveform (Lamanna et al. 2015; Losilla et al. 2020).

We recently performed an examination of differential gene expression among closely related members of the genus *Paramormyrops* (Losilla et al. 2020) where we identified expression correlates of EOD duration. Based on the results of this study, we were motivated to link observed patterns of differential gene expression between species to mechanisms underlying changes in EOD duration that occur within the lifetime of an individual. Previous studies have indicated that androgen hormones, including 17 α -methyltestosterone (17 α MT) increase EOD duration in several mormyrid species, an effect that mimics natural sex differences (Bass and Hopkins 1983, 1985; Bass et al. 1986; Bass and Volman 1987; Freedman et al. 1989; Herfeld and Moller 1998). Although elongated EODs during the breeding season are a male character, androgen hormone treatment of female mormyrids elicits a male-like response in the duration of the EOD (Bass and Hopkins 1983, 1985; Bass et al. 1986; Bass and Volman 1987; Freedman et al. 1989; Herfeld and Moller 1998). Androgen hormones provoke similar responses on the duration of male and female EODs in gymnotiforms (e.g. (Mills and Zakon 1987)).

Treatment with androgen hormones can provoke large changes in EOD duration under controlled circumstances, providing a hitherto unexplored opportunity to identify the expression correlates of EOD duration differences. In this study, we leveraged this paradigm to investigate the molecular underpinnings of changes in EOD duration in the mormyrid *Brienomyrus brachyistius*, a species where males greatly elongate their EOD under breeding conditions, influenced by their social status and in connection with androgen levels in blood (Carlson et al. 2000). We experimentally applied 17 α MT to *B. brachyistius* individuals over an 9-day experimental period, and examined patterns of differential gene expression over the course of this treatment compared to control organisms. Unlike our previous study, which compared gene expression between species, this analysis focused on

conspecific adult individuals in controlled environments, eliminating critical confounding factors like phylogenetic divergence and additional EOD variation. Our analysis strongly supports known aspects of morphological and physiological bases of EOD duration, and for the first time identifies specific genes and broad cellular processes that alter morphological and physiological properties of electrocytes during seasonally plastic EOD changes, most striking among these is the differential expression of multiple K⁺ voltage-gated channels.

Methods

Study fish

We drew 21 *Brienomyrus brachyistius* individuals (20 survived) from our laboratory colony, purchased through the pet trade. We limited our experiment to fish deemed adults (standard length \geq 110 mm); and to the best of our abilities, to males, identified by the presence of an indentation in the anal fin, a common dimorphic feature in mormyrids. Although we expected similar responses to androgen treatment from adult males and females, the latter have a higher intrinsic value in our colony.

Experimental conditions

During the experiment, we housed each fish individually in 38 L tanks filled with 30 L of water and equipped with one mechanical and biological filter. Tank dimensions were 25 cm width x 50 cm depth x 30 cm height. The front and back ends of the tank correspond to the 25 cm walls. We kept the fishes under 12 h light cycles, fed them blackworms daily, monitored levels of ammonia, nitrites, and nitrates, and performed 30% water changes as needed during the acclimation periods (no water changes were carried out after experimental treatments were applied). The following per tank water conditions were monitored daily or every other day and remained within the indicated limits: conductivity 500-570 μ S/cm, temperature 24-26 °C, pH 7.0-7.5. Each tank was provided with a custom-built housing arrangement that served as a fish shelter but also facilitated measuring EODs with the fish at a stable location. These arrangements consisted of a PVC tube attached to an aquarium plastic egg crate via two zip ties, the egg crate was cut to the width of the tank and fixed to the bottom with four

suction cups with hose clips, each with a short piece of PVC tube acting as a lock. The housing tube laid centered at the bottom, oriented from the front to the back end of the tank (Fig. 2.2).

EOD recordings

We recorded EODs from each fish on a daily basis during acclimation and experimental periods (median 19 EODs per fish/day, range 14-20). During recording sessions, electrodes were attached to suction cups with hose clips affixed to the center of the front (positive electrode), back (negative), and left (neutral) sides of each tank (Fig. 2.2). The distance between the positive and negative electrodes was 48 cm. Signals were recorded with bipolar chloride-coated silver wire electrodes, amplified (bandwidth = 0.0001-50 kHz, gain = 20) with a bioamplifier (model BMA-200; CWE, Inc), digitized using a multifunction DAQ device (model USB-6216 BNC; NI) and the recording software EODrecorder2 (Terminal Configuration = Differential, sampling rate = 250 KHz, voltage range = +/- 5 V, npts = 2048). On each EOD, we used custom code to correct baseline noise, normalize by its peak-to-peak amplitude, center time = 0 at P1, and determine the start and stop points (and hence the duration).

Experimental treatments

Each fish was allowed to acclimate to the housing tank for a minimum of two weeks before the start of a treatment. During this time a subject's EODs were recorded daily and the coefficient of variation (CV) of the daily mean EOD duration was calculated for the last five days (including the current day). We deemed a fish acclimated when the CV was < 5% for five consecutive days.

We carried out our experimental treatments in two back-to-back cycles, with nine (eight survived) and 12 fish, respectively. For each cycle, we call day 0 the day of treatment application, administered on all fish of the cycle, immediately after taking the EOD recordings for that day. Fish were assigned to one of three experimental treatments, and each cycle started with equal numbers of fish from each treatment. Fish in two treatments, "T8day" and "T1day", received 3 mg/L of 17 α MT dissolved in 5ml of ethanol 100%; whereas fish in the third treatment, "control", received 5ml of ethanol 100%. In all cases, ethanol was added to the fish tank water. We monitored the fish and recorded EODs daily,

until and including the assigned dissection date. This date was on day 1 for treatment T1day, and on day 8 for treatments T8day and control. The 20 study individuals were apportioned seven to each of the treatments T8day and T1day, and six to the control treatment.

Our experimental setup (fixed housing arrangement, EOD recordings taken in each fish's tank, and non-invasive 17 α MT addition) was designed to minimize fish manipulation and disturbance. The study individuals did not undergo surgery or injections, and once placed in its experimental tank, a fish was not removed or otherwise handled until its dissection date. All procedures complied with federal and state regulations and were approved by Michigan State University's Office of Environmental Health & Safety, and Institutional Animal Care & Use Committee.

We tested for differences in EOD duration with a one-way ANOVA at 1) day 0, and 2) the last day of treatment (i.e. day 1 for T1day, day 8 for control and T8day). Where significant results were detected, we examined differences in all pairwise treatment comparisons with the post-hoc Tukey HSD test.

Tissue dissections, RNA extraction and sequencing

Individual specimens were euthanized with an overdose of MS-222 (0.7 g MS-222 in 2 L of fish tank water; and 1.4 g of sodium bicarbonate to buffer pH and thus reduce fish discomfort). We skinned and dissected caudal peduncles and stored them in RNA-later (Ambion, Inc) following manufacturer's instructions until processing. Sex was confirmed by gonad inspection after which we believe that 19 fish were males and one was female.

All RNA extraction, library preparation, and sequencing were performed by Genewiz, LLC (South Plainfield, NJ). Total RNA was extracted using Qiagen RNeasy Plus Universal mini kit following manufacturer's instructions (Qiagen, Hilden, Germany). Extracted RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked using Agilent TapeStation 4200 (Agilent Technologies, Palo Alto, CA, USA). RNA poly-A selected sequencing libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina using manufacturer's instructions (NEB, Ipswich, MA, USA). Briefly, mRNAs were first enriched with Oligo(dT) beads. Enriched

mRNAs were fragmented for 15 minutes at 94 °C. First strand and second strand cDNAs were subsequently synthesized. cDNA fragments were end repaired and adenylated at 3'ends, and universal adapters were ligated to cDNA fragments, followed by index addition and library enrichment by limited-cycle PCR. The sequencing libraries were validated on the Agilent TapeStation and quantified using Qubit 2.0 Fluorometer as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, USA). The sequencing libraries were clustered on flow cells. After clustering, the flow cells were loaded on to the Illumina HiSeq instrument (4000 or equivalent) according to manufacturer's instructions. The samples were sequenced using a 2x150bp Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

Read processing and data exploration

We inspected raw and processed reads with FastQC v0.11.7 (Babraham Bioinformatics) and used Trimmomatic v0.39 (Bolger et al. 2014) to remove library adaptors, low quality reads, and filter small reads. The succeeding steps were executed using scripts included with Trinity v2.11.0 (Grabherr et al. 2011; Haas et al. 2013). We aligned reads from each specimen to the predicted transcripts of the Maker v2.31.10 (Holt and Yandell 2011) annotation of our *Brienomyrus brachyistius* genome v0.3 (unpublished), with bowtie2 v2.3.4.1 (Langmead and Salzberg 2012). Expression quantification was estimated at the gene level using RSEM v1.3.0 (Li and Dewey 2011), followed by exploration of the data with a gene expression correlation matrix based on Euclidean distances and Pearson's correlation coefficient (for genes with read counts > 10, Trinity's default parameters).

Differential Gene Expression

We identified differentially expressed genes (DEG) between experimental treatments in the three possible pairwise comparisons (control vs T8day, control vs T1day, T1day vs T8day) with edgeR v3.20.9 (Robinson and Oshlack 2010) through a script provided with Trinity. For each comparison

(contrast), we conservatively identified differentially expressed genes (significant DEG) as those genes with a minimum expression fold change (FC) of 4 and p-value < 0.001 after FDR correction.

Functional enrichment analysis

We employed mitch v1.0.10 (Kaspi and Ziemann 2020) to detect sets of genes that exhibit joint up- or downregulation across our three contrasts. As inputs, mitch needs a gene set library and profiled expression data. The latter was the whole set (i.e. not filtered by FC or p-value thresholds) of edgeR differential expression results from each contrast. For the gene set library, we started with the *Danio rerio* gmt file with gene ontology (GO) (Ashburner et al. 2000) terms as gene sets, available from <http://ge-lab.org/gskb/>. This file includes GO terms from the three GO domains (Cellular Component, Biological Process, and Molecular Function). We split this file into three, one for each GO domain. Preliminary results using each of these files as input were redundant, thus we limit our analysis to the GO domain Biological Process, which contained the most gene sets.

Since the gene identifiers in the gmt file are different than those of the edgeR results, mitch requires a third input file that relates these gene identifiers. We generated this file with custom code, from the gff3 annotation file produced by Maker. Unannotated *B. brachyistius* genes were excluded from the analysis. Also excluded were gene sets with fewer than five genes in the edgeR results (minsetsize option in mitch_calc function). We filtered mitch's enrichment result to gene sets with FDR-corrected p-value < 0.01 and the higher dimensional enrichment score (S) > 0.1, to minimize false positives. After filtering, the results of prioritizing by "significance" or "effect" (priority option in mitch_calc function) were identical.

To further simplify the analysis, we subjectively grouped the gene sets identified by mitch into broad functional categories, inferred their expression patterns across the experimental treatments, and commented on how these patterns may relate to the hormonal manipulations applied and the observed changes in EOD duration.

Differentially expressed genes of highest interest for EOD duration

We used two procedures to identify the most interesting genes in relation to EOD duration from the significant DEG detected by edgeR. First, we take a deeper look at the gene sets in the broad categories of most interest from the previous step. We compared the genes that make up each of these gene sets with the significant DEG from each contrast, and highlight the genes common to both sources. Our second procedure consisted of manually selecting additional significant DEG from each contrast based on their annotation, guided by general functional themes suspected to affect EOD phenotype. Finally, we remark how expression changes in several of these select genes are likely to affect the duration of the EOD.

We found no outliers for gene expression or EOD duration, hence we analyzed samples lumped together in regard to sex and experimental cycle. The code we used in this work can be found at <https://github.com/msuefishlab/>.

Results

EOD duration responses to hormone treatment

There were no differences in the duration of the EOD between the treatments on day 0, immediately before treatment application ($p = 0.15$, Fig. 2.3). On the last day of treatment, shortly before tissue dissection, the duration of the EOD differed between the treatments ($p < 0.001$); specifically, EOD duration increased in T8day compared to control ($p < 0.001$) and to T1day ($p < 0.001$), but it did not differ between control and T1day ($p = 0.21$) (Fig. 2.3). Throughout the experiment, we observed mean EOD duration values stable during the last days of the acclimation period for all treatments and also during the experimental period for treatments control and T1day; however there was a sharp increase in mean EOD duration for fish in the T8day treatment upon addition of 17α MT (Fig. 2.4). We provide a graphical illustration of changes in EOD duration in a representative fish from each treatment in Fig. 2.5.

RNAseq data and differentially expressed genes

We explored the RNAseq data with a heatmap of pairwise correlations of gene expression across all 20 samples. Gene expression patterns across all specimens were highly correlated (Pearson's $r > 0.926$), and correlation values were higher between fish from the same treatment (Fig. 2.6). At the level of treatments, these correlations indicate that the treatments T1day and T8day have a more similar gene expression pattern than with control.

We performed the three possible pairwise DGE comparisons between treatments. They each found a similar number of expressed genes, close to 21400; with only a small percentage differentially expressed (range 0.18 - 1.18). Among these DEG ($FC > 4$, FDR-corrected p -value < 0.001), there were always more genes upregulated in the treatment that entailed longer exposures to 17 α MT (Table 2.1).

Functional enrichment analysis

This analysis compares patterns of expression of pre-defined groups of genes (gene sets, here composed of the GO terms for the GO domain Biological Process) across our three pairwise treatment comparisons (contrasts). The level of enrichment (= up- or downregulation) of each gene set in each contrast is quantified by an enrichment score (s , range -1 to 1). These contrasts were constructed such that positive values of s represent upregulation with increased exposure to 17 α MT, and negative values of s indicate downregulation with increased exposure to 17 α MT.

Our experimental design facilitates inferences about gene expression over the amount of time subjects were exposed to 17 α MT (Fig. 2.7). The comparison control vs T8day informs about broad changes in gene expression over the course of the experiment, whereas control vs T1day exposes early changes in gene expression, and T1day vs T8day uncovers late changes in gene expression. For example, a hypothetical gene set may show little 'broad' changes (control vs. T8day), but important 'granular' changes in opposite directions before and after day 1 (control vs T1day, and T1day vs T8day). We note that we never observed conflicts between 'broad' and 'granular' expression patterns.

We detected 39 gene sets enriched across the three contrasts studied (Fig. 2.8). To aid with our interpretation, we further grouped these gene sets into seven general categories: “energy metabolism”, “gene expression”, “signal transduction”, “protein transport”, “structural/morphology”, “cation homeostasis”, and “too broad” (Fig. 2.8). In the latter we binned gene sets too general to assign to a given category. Although a subjective sorting, this classification greatly facilitates the interpretation of the overall response in gene expression from the electrocytes to our experimental manipulations.

Gene sets in the category “energy metabolism” were upregulated over time and they displayed, overall, the highest expression increases (i.e. Fig. 2.8, cluster 1). In the category “gene expression”, gene sets showed broad upregulation (control vs. T8day), but granular changes reveal that upregulation happened quickly (control vs T1day), whereas later changes, (T1day vs T8day), were of smaller magnitude, and often involved downregulation (i.e. Fig. 2.8, cluster 3). A similar enrichment pattern was observed in the “protein transport” category, with the slight difference that there was virtually no change between days 1 & 8 (i.e. Fig. 2.8, clusters 2 & 3). Gene sets classified as “signal transduction” were consistently downregulated as the experiment progressed, although the largest downregulations occurred by day 1 (i.e. Fig. 2.8, cluster 4). Finally, the gene sets in the categories of “cation homeostasis” and “structural/morphology” exhibited variable enrichment directions in the broad control vs T8day contrast; however, the granular changes uncovered downregulation (or little change) at day 1, followed by a change in the enrichment direction, manifested as upregulation (or little change) at day 8 relative to day 1 (i.e. Fig. 2.8, clusters 2 & 4). This pattern of gene expression over the time course of 17aMT exposure suggests that gene expression changes for these two gene sets are not simply the immediate consequences of the hormone treatment, but rather downstream responses. Because of our overall motivation to examine genes that could contribute to anatomical and physiological changes that would affect EOD duration, we focused on these two categories. “Cation homeostasis” contains the gene sets “calcium ion transmembrane transport”, “ion transmembrane transport”, and “ion transport”; and

“structural/morphology” comprises the gene sets “adhesion” and “microtubule-based process”. For the remainder of our analysis, we refer to these five gene sets as “key gene sets”.

Differentially expressed genes of highest interest for EOD duration

We generated a list of genes that may directly influence EOD duration by selecting all genes common to at least one “key gene set” and at least one of the three lists of DEG from the pairwise comparisons (Table 2.2). This produced 14 genes, none came from the gene set “calcium ion transmembrane transport”. We discarded gene *cox6a1*, its product participates in the mitochondrial respiratory chain and therefore its function is related to “energy metabolism”. In addition, we are suspicious of genes *ninj1* and *kcnq1*: inspection of their per-sample CPM-normalized gene counts revealed very low levels of expression for both genes across all treatments (*ninj1*: no expression in control, median expression in T1day = 0, in T8day = 0.85; *kcnq1*: no expression in T1day, median expression in control = 1.03, in T8day = 0).

Finally, we built an additional list of genes by subjectively selecting the remaining DEG from the pairwise comparisons those that, based on their annotation, could contribute to EOD duration. As a guide, we used the general, functional themes suspected to affect EOD phenotype (Gallant et al. 2012, 2014; Lamanna et al. 2015; Losilla et al. 2020). For simplicity, we use the same themes we have used previously (Losilla et al. 2020): “extracellular matrix”, “cation homeostasis”, “lipid metabolism”, and “cytoskeletal & sarcomeric”. This list contains 36 genes, their breakdown by general theme and featured examples are provided in Table 2.3.

Discussion

In this study, we sought to identify gene expression patterns related to changes in EOD duration in samples from the same species, *Brienomyrus brachyistius*, thus reducing confounding sources of variability. To do so, we experimentally treated individually housed specimens with 17 α MT for 1 or 8 days and compared patterns of gene expression to individuals treated with ethanol vehicle. By

leveraging on differential expression and functional enrichment analyses, we were able to systematically identify specific genes and broad functional themes that changed expression as the EOD elongated.

EOD Duration and gene expression in the electric organ

Unlike other features of EODs like complexity or polarity, duration can change relatively quickly depending on conditions. Therefore, we housed fish individually, carefully monitored their acclimation, minimized disturbances, and elicited EOD elongation with a proven hormone manipulation. The resulting EODs were consistent through time and of similar duration across treatments before addition of 17 α MT, and upon its application the recipient fish steadily increased their EODs (Fig. 2.3, Fig. 2.4, & Fig. 2.5). This response was not evident by day 1, but it increased progressively and became striking by day 8. The critical EOD duration comparison between treatments is that of the last day of treatment (Fig. 2.3, right pane) because the underlying recordings were taken shortly before tissue dissection. In other words, all the gene expression data derives from the electrocytes that generated these and only these EODs.

Likewise, our experimental design addressed sources of unwanted genetic variability. All our samples came from adults of the same species, *Brienomyrus brachyistius*, thus reducing differences in gene expression attributable to phylogenetic divergence and life cycle stage. We think that this background genetic homogeneity contributed to the low levels of differential expression detected (<1.2%, Table 2.1) and to the resulting high correlation levels of gene expression across all 20 samples. Importantly, these correlation values were higher among samples of the same treatment (Fig. 2.6), implying that the differences in gene expression are driven by experimental manipulation. Notably, at treatment level, overall gene expression was more similar between T1day and T8day (Figs. 6 & 8, Table 2.1), despite the seven day difference between them, compared to a one day difference between treatments control and T1day. This strongly suggests that the majority of changes in gene expression had materialized by day 1, despite little change in EOD duration at this time (Fig. 2.3, Fig. 2.4, & Fig. 2.5). This exalts the relevance of treatment T1day: it provides granularity to the understanding of how 17 α MT

induces EOD elongation (Fig. 2.7). In sum, given the robustness of our EOD duration and gene expression data, we think that it is very reasonable to expect that gene expression changes that underlie differences in EOD duration are amongst the DEG we detected.

Functional enrichment analysis

Functional enrichment analysis greatly simplifies the interpretation of the copious amount of data generated by omic technologies like RNA sequencing, by identifying previously defined groups of related genes that differ between the experimental treatments (García-Campos et al. 2015). We employed mitch, a multi-contrast enrichment analysis tool, to detect gene ontology terms of biological processes jointly enriched across our three pairwise treatment contrasts (Fig. 2.8). To facilitate inferences about how these results may relate to our experimental manipulations and the ensuing changes in EOD duration, we further group these significant gene sets into broad categories (Fig. 2.8). The per-category expression patterns uncovered a broad but coherent picture of sequential changes in gene expression that presumably start with 17 α MT-induced modifications and likely culminate in a distinct set of gene expression changes that alter the EOD phenotype.

We note that this analysis is limited by errors in the *Danio rerio* gmt file and in the *B. brachyistius* genome annotation, which cause the unavoidable exclusion of genes from this procedure. Despite these shortcomings, the results from our functional enrichment analysis delineate very coherent steps of gene expression changes that culminate with gene sets indicative of the usual functional suspects of underscoring EOD phenotypic variation.

Early Changes in Response to 17 α MT

Upon hormone-hormone receptor binding, cell signaling pathways transduce the signal to downstream effectors. As part of the response, changes in gene expression, and increases in protein transport and in energy consumption are expected, regardless of the ultimate, phenotypic effects induced. Accordingly, we observed very fast enrichment changes in the categories of "signal transduction", "gene expression", "protein transport", and "energy metabolism" (Fig. 2.8, control vs

T1day contrast). After day 1, the enrichment change in these categories is more subdued (except in "energy metabolism"), suggesting that they mainly represent an immediate or general reaction to hormone exposure (Fig. 2.8, T1day vs T8day comparison). We propose that these immediate responses ultimately induce expression genes in the categories structural/morphology and cation homeostasis, and that genes in these are the effectors of the observed EOD elongation.

Late Changes in Response to 17 α MT

"Energy metabolism", on the other hand, consistently displays very strong upregulation throughout the experiment, including the time when EOD duration has palpably increased. Suitably, longer EODs are predicted to be metabolically more costly (Hopkins 1999a), and this has been corroborated in the gymnotiform *Brachyhypopomus gauderio* (Salazar and Stoddard 2008). In addition, genes in the categories "structural/morphology" and "cation homeostasis" display marked expression changes after day 1 (Fig. 2.8), and they embody the major features thought to be responsible for EOD variation. Genes in the category "cation homeostasis" could alter the duration of the EOD by regulating plasma membrane ion channels, as has been demonstrated in gymnotiforms (reviewed in (Markham 2013; Dunlap et al. 2017)) and has been explored in mormyrids (Lamanna et al. 2015; Nagel et al. 2017; Swapna et al. 2018). Genes in the category "structural/morphology" may underlie changes in the surface area of the electrocyte membrane, a character that has been correlated with longer EODs in descriptive (Bennett 1971; Paul et al. 2015; Nguyen et al. 2020) and 17 α MT-mediated experimental work (Bass et al. 1986; Bass and Volman 1987; Freedman et al. 1989). Also included in this category are the genes that may regulate electrocyte interactions with the extracellular matrix (ECM) and the connective tissue sheath, which may contribute to EOD phenotype (Losilla et al. 2020). Noteworthy, signal transduction through Rho GTPases (Fig. 2.8) is a key regulatory mechanism of the actin cytoskeleton, membrane protrusions, and adhesion complexes (Hall 1998).

Differentially expressed genes of highest interest for EOD duration

We condensed our results to a set of genes most likely to modulate EOD duration by 1) taking the genes common to the DGE analysis and to the key gene sets from the functional enrichment analysis (Table 2.2); and, given the limitations of the functional enrichment procedure, 2) by further selecting DEG that align with previous knowledge (Table 2.3). One bias of the latter is our inevitably incomplete understanding of the functional effects of every gene. Unless otherwise indicated, gene descriptions come from Uniprot (The UniProt Consortium 2019).

Genes that affect electrocyte morphology

The anterior and posterior faces of the electrocyte plasma membrane often increase their surface area by evaginations and invaginations of the membrane (Schwartz et al. 1975). These projections, depending on their characteristics, are referred to in the literature as *papillae*, *tubules*, *calveoli*, or *canaliculi*. Increases in the membrane surface area are associated with longer EODs (Bennett 1971; Paul et al. 2015), mainly through projections on the anterior face (Bass et al. 1986; Freedman et al. 1989; Paul et al. 2015; Nguyen et al. 2020). Purportedly these increases in membrane surface area increase membrane capacitance that thus delay spike initiation (Bass et al. 1986; Bass and Volman 1987). At least the larger membrane projections are supported by the cytoskeleton (Kornienko et al. 2021). Given this framework, we predict cytoskeletal & sarcomeric genes involved in longer EODs to be upregulated in the treatment T8day, and indeed every highlighted gene from this theme displays this expression pattern: *tubb4b* (Table 2.2), *act2*, *actn1*, *eml6*, *tubb2a*, and *xirp1* (Table 2.3). Furthermore, we expect that this increase in surface area also requires rearrangements in the phospholipid bilayer and in the electrocyte's interactions with the ECM and the connective tissue sheath. Previous research supports that these themes are related to variation in EOD phenotypes (lipid metabolism (Lamanna et al. 2015; Losilla et al. 2020), ECM (Losilla et al. 2020)). In fish with longer EODs, we identified increased gene expression of two paralogs of the gene *fam126a* (plays a key role in certain cell types with expanded plasma membrane), *murc* (modulates the morphology of existing membrane invaginations),

and the adhesion genes *cdh11*, *edil3*, and *thbs4b*; whereas we found reduced expression of *epd12* (Table 2.2 & Table 2.3). We have previously identified *epd12*, whose product may participate in cell-matrix adhesion, as a gene that could affect the EOD waveform (Losilla et al. 2020) and it is the focus of ongoing work.

However, it is unlikely that these morphological changes alone are sufficient to explain the range of variation observed in the duration of mormyrid EODs. For example, (Korniienko et al. 2021) showed that increases in membrane surface area do not linearly correlate with EOD duration increases in *Campylomormyrus rhynchophorus* during development. This can be modeled simply by summing two identical action potentials with opposite sign and systematically varying their relative delay (Fig. 2.9B): if larger surface areas extend EODs by delaying depolarization, this model demonstrates diminishing effects on EOD duration as the relative delay of APs firing increases between the posterior and anterior faces (Fig. 2.9A, bottom). Amplitude, on the other hand, would increase constantly across this range (Fig. 2.9A, top). This observation is also supported by more sophisticated computer simulations based on the EOD of the gymnotiform *Steatogenys elegans*, where longer firing delays between the APs from the posterior and anterior faces produced increases in EOD amplitude but much more subdued rises in EOD duration (Markham and Zakon 2014).

Genes that affect electrocyte physiology

Another mechanism to change EOD duration is to alter the electrocyte ionic currents, which can be achieved by modifying the functional properties or levels of expression of ion channels. In gymnotiforms, ionic currents are largely driven by voltage-gated Na⁺ and K⁺ currents (e.g. (Shenkel and Sigworth 1991; Ferrari and Zakon 1993)) and their modulation via androgen hormones affects EOD duration (e.g. (Zakon et al. 1991; Ferrari et al. 1995)). To the best of our knowledge, technical difficulties have hindered recordings of mormyrid electrocyte ionic currents, yet significant convergent evolution is expected. It has been proven that both Na⁺ (Zakon et al. 2006; Arnegard et al. 2010b) and K⁺ (Nagel et al. 2017; Swapna et al. 2018) voltage-gated channels are expressed in mormyrid electrocytes, and

differential expression of ion channel genes has been linked with EOD variation (Lamanna et al. 2015; Losilla et al. 2020).

First, we identified a Na⁺ channel beta subunit, *scn4b*, downregulated in samples with longer EODs (Table 2.3). This perfectly mirrors results observed in the gymnotiform *Sternopygus macrurus*, where exposure to an androgen hormone elongated the EOD and downregulated the expression of a Na⁺ channel β 1-subunit, that probably speeds inactivation rates of the α subunits (Liu et al. 2007). Given this mode of action, suppression of the β subunit would likely extend the electrocyte Na⁺ currents and the EOD.

Second, we observe a striking coordinated regulation of at least seven K⁺ voltage-gated channels: *kcna1*, *kcng4*, three paralogs of *kcnc2* (Table 2.2); *kcnh7* (Table 2.3); and *kcna7* (BBRACH_0.3_G032307, K⁺ voltage-gated channel subfamily A member 7, see below). The identities of these genes will be revisited when more exhaustive genomic resources for *B. brachyistius* become available, nevertheless we expect that their functional grouping will not change.

Based on the model of EOD production (Fig. 2.1), APs from each electrocyte face, and therefore the resulting EOD, would elongate if K⁺ efflux is delayed or its rate reduced. Allowing K⁺ ions out of the cell is the canonical function of delayed-rectifier K⁺ channels; and two of these, Kv1.1a and Kv1.2a, are less expressed in longer EODs in *S. macrurus*, a response that can be elicited by androgen hormones (Few and Zakon 2007). We detected five differentially expressed delayed-rectifier K⁺ channels (six if the suspicious *kcng1* [Kv7.1] is included), and in perfect agreement with expectations, all of them were strongly downregulated in fish with longer EODs (T8day treatment): *kcna1* (Kv1.1), *kcna7* (Kv1.7), and three paralogs of *kcnc2* (Kv3.2). Interestingly, Kv3.2 channels show adaptations for fast repolarization in neurons (Rudy and McBain 2001; Lien and Jonas 2003), positioning these paralogs candidates for electrocyte subfunctionalization in EODs of short duration. We also identified gene *kcng4* (Kv6.4), but unlike the previous K⁺ voltage-gated channels genes, this one was found upregulated in longer EODs.

Nevertheless, the expectation that the net effect is to delay or reduce the rate of K⁺ efflux is probably still met: the protein coded by this gene is classified as a modifier/silencer channel that modulates the delayed-rectifier K⁺ channel *kcnb1* (Kv2.1), by increasing its slow time constant of inactivation (Mederos y Schnitzler et al. 2009). What this means is that, when associated with *kcng4*, the delayed-rectifier *kcnb1* takes longer to inactivate, extending the electrocyte's APs (Fig. 2.1) and the resulting EOD. Finally, the inwardly-rectifying K⁺ voltage-gated channel *kcnh7* (Kv11.3) was also detected at higher levels in longer EODs. These channels mainly import K⁺ into the cell, which helps restore the membrane resting potential after the AP. If longer EODs expel more K⁺ ions, a reasonable possibility given their longer underlying APs, then the upregulation of *kcnh7* is not surprising. However, this indicates that the elevated expression of this gene is more likely a consequence than a cause of longer EODs.

Finally, we note that *kcna7* was not detected in our differential expression analysis; however, it has been reported that this gene codes for mormyrid-exclusive amino acid substitutions that enable ultra-brief EODs and that it is expressed at much higher levels than other K⁺ voltage-gated channels in mormyrid electrocytes (Swapna et al. 2018). Upon checking its expression values in our data, it was indeed expressed at much higher levels than all related genes we studied, and it was downregulated in T8day fish in the two comparisons control vs T8day and T1day vs T8day (in each case FC = -3.0 and FDR-corrected p-values = 0.007 and < 0.001, respectively). Given this, we include this gene in our discussion and suggest that the significance threshold for differential expression of FC = 4 may be a too stringent for ion channel genes. This highlights how remarkable is the number of K⁺ voltage-gated channels we detected; but we note that, given its larger expression values, *kcna7* may be the principal contributor to EOD elongation.

Taken together, our K⁺ channel results perfectly agree with each other, with the model of EOD function (Fig. 2.1), and with observations in gymnotiforms (Few and Zakon 2007). They are, however, somewhat at odds with part of the results from a recent study on two *Campylomormyrus* species, where

11/13 *kcna* genes (delayed-rectifiers) showed at least a tendency for higher expression in EOs of the species with longer EODs (Nagel et al. 2017). Although the only gene common with our study is *kcna1*, based on our results we would expect delayed-rectifying K⁺ voltage-gated channels to be generally downregulated in longer EODs. Factors different from EOD duration, like intrinsic gene expression differences between their target species, or their very dissimilar waveform shapes, may underlie the discrepancies. Alternatively, the delayed-rectifier K⁺ voltage-gated channels they studied may behave differently, for example they could be modulated by other proteins that increase their activation and inactivation times.

Conclusions

We designed this study with the objective of detecting gene expression changes driven by the elongation of the EOD in mormyrid fish. As such, we took steps to minimize confounding factors that could influence gene expression. For example, experimental fish were housed individually, given ample time to acclimate, and fish manipulation and disturbances were kept to a minimum, by eliminating surgical procedures, injections, and the need to record EODs in an external tank. In addition, we affixed the housing tube and the electrode positions to facilitate standardized recording conditions. The most critical factor we controlled for is species: by focusing on adults of one species, *Brienomyrus brachyistius*, we avoided gene expression variation resulting from interspecific differences, life stage, and additional variable characters of the EOD waveform. We believe that this work represents the best controlled effort to date to study changes in gene expression that stem from changes in an EOD character. A recommendation for future work is to reduce the concentration of 17 α MT to 2 mg/L, with a proportional reduction in the ethanol vehicle. It is possible that ethanol briefly affected bacteria in the fish tanks, manifested as transitory water clouding observed in some cases, without detectable effects on the fishes' behavior or EOD.

Our results provide insights into the stepwise cellular process induced by androgen hormones and that lead to EOD elongation. A fast, coordinated response, mediated through Rho and other small G-proteins, activates gene expression, protein transport and processes that increase energy availability. The downstream result is changes in gene expression that modify the morphological and physiological properties of electrocytes, in ways that align perfectly with theoretical expectations on how these changes could increase EOD duration. Increases in electrocyte membrane surface area are associated with longer EODs, and we detected a concordant upregulation of genes that participate in plasma membrane expansions. Importantly, our results deepen our understanding of how an increase in membrane area is accomplished, by pointing out that these morphological changes likely occur in tandem with changes to electrocyte interactions with the ECM and to the supporting cytoskeleton, including increases in specific actin and tubulin proteins, two of its essential building blocks. Similarly, our results also support that changes to the physiological properties of the electrocyte, mediated by expression changes of ion channels, influence EOD duration. They implicate changes to ionic currents that seamlessly agree with current theories of EOD function and are novel examples of convergence between gymnotiforms and mormyrids. These are the downregulation in longer EODs of a Na⁺ channel beta subunit and delayed-rectifier K⁺ channels. K⁺ voltage-gated channels embody our most striking result, because they reveal that the expression of a wide range of these channels, centered in but not limited to delayed-rectifiers, is consistently modulated in accordance to expectations of how they would influence EOD duration. Our results not only support and deepen current ideas of how mormyrids modulate EOD duration, by they also provide a list of specific genes likely to mediate this regulation. Given the central role of EOD duration to EOD diversification, and the importance of the EOD in species divergence, these genes may be significant participants in the speciation process amid African weakly electric fish.

APPENDIX

Table 2.1. All three possible pairwise differential gene expression comparisons with the number of genes expressed, differentially expressed, and their breakdown by treatment

comparison	genes expressed	genes differentially expressed (%)	upregulated genes	
			treatment	total
control vs T8day	21466	254 (1.18)	control	103
			T8day	151
control vs T1day	21622	97 (0.45)	control	38
			T1day	59
T1day vs T8day	21121	37 (0.18)	T1day	12
			T8day	25

Table 2.2. Genes differentially expressed in the three pairwise treatment contrasts that also belong to the key gene sets

gene set	gene name	code in <i>B. brachyistius</i> genome	annotation	contrast	FC
adhesion	<i>cdh11</i>	BBRACH_0.3_G034977	Cadherin-11	control vs T8day	5.1
adhesion	<i>edil3</i>	BBRACH_0.3_G026692	EGF-like repeat and discoidin I-like domain-containing protein 3	control vs T8day	5.1
adhesion	<i>ninj1</i>	BBRACH_0.3_G046208	Ninjurin-1	control vs T8day	186.8*
adhesion	<i>thbs4b</i>	BBRACH_0.3_G021564	Thrombospondin-4-B	control vs T8day	33.6
				control vs T1day	5.9
ion transmembrane transport, ion transport	<i>gabrb3</i>	BBRACH_0.3_G024733	Gamma-aminobutyric acid receptor subunit beta-3	control vs T8day	7.1
				T1day vs T8day	8.2
ion transport	<i>kcna1</i>	BBRACH_0.3_G023553	Potassium voltage-gated channel subfamily A member 1	control vs T8day	-35.7
ion transport	<i>kcnc2</i>	BBRACH_0.3_G012974	Potassium voltage-gated channel subfamily C member 2	control vs T8day	-22.1
				control vs T1day	-4.8
				T1day vs T8day	-4.6
ion transport	<i>kcnc2</i>	BBRACH_0.3_G012975	Potassium voltage-gated channel subfamily C member 2	control vs T8day	-21.2
				control vs T1day	-4.0
				T1day vs T8day	-5.2
ion transport	<i>kcnc2</i>	BBRACH_0.3_G052206	Potassium voltage-gated channel subfamily C member 2	control vs T8day	-12.0
ion transport	<i>kcng4</i>	BBRACH_0.3_G032339	Potassium voltage-gated channel subfamily G member 4	control vs T8day	8.0
ion transport	<i>kcng1</i>	BBRACH_0.3_G007081	Potassium voltage-gated channel subfamily KQT member 1	control vs T1day	-206.0*
ion transport	<i>trpv1</i>	BBRACH_0.3_G003271	Transient receptor potential cation channel subfamily V member 1	control vs T8day	4.8
				control vs T1day	5.8
microtubule-based process	<i>tubb4b</i>	BBRACH_0.3_G013797	Tubulin beta-4B chain	T1day vs T8day	4.7

FC = fold change in gene expression; positive values represent upregulation, and negative values indicate downregulation, in the treatment that involved longer exposures to 17 α MT

* Genes with very low expression across all treatments. See main text for details

Table 2.3. Select differentially expressed genes from the three pairwise treatment contrasts. Shown are the number of genes in each general theme and featured gene examples from each general theme

general theme	total	featured examples				
		gene name	code in <i>B. brachyistius</i> genome	annotation	contrast	FC
cation homeostasis	7	<i>kcnh7</i>	BBRACH_0.3_G007175	Potassium voltage-gated channel subfamily H member 7	control vs T8day	296.6
					T1day vs T8day	296.4
		<i>scn4b</i>	BBRACH_0.3_G059092	Scn4b: Sodium channel subunit beta-4	control vs T8day	-9.5
Cytoskeletal & sarcomeric	15	<i>act2</i>	BBRACH_0.3_G042661	Actin, alpha skeletal muscle 2	control vs T1day	-5.4
					T1day vs T8day	8.1
		<i>actn1</i>	BBRACH_0.3_G018329	Alpha-actinin-1	control vs T8day	4.9
		<i>eml6</i>	BBRACH_0.3_G034764	Echinoderm microtubule-associated protein-like 6	control vs T8day	5.0
		<i>tubb2a</i>	BBRACH_0.3_G022786	Tubulin beta-2A chain	control vs T8day	4.1
		<i>xirp1</i>	BBRACH_0.3_G014353	Xin actin-binding repeat-containing protein 1	control vs T1day	46.1
					control vs T8day	54.5
Extracellular matrix	6	<i>epdl2</i>	BBRACH_0.3_G057682	Ependymin	control vs T8day	-5.2
Lipid metabolism	8	<i>fam126a</i>	BBRACH_0.3_G055176	Hyccin	control vs T1day	6.1
					control vs T8day	5.3
		<i>fam126a</i>	BBRACH_0.3_G058365	Hyccin	control vs T1day	7.3
					control vs T8day	6.8
		<i>murc</i>	BBRACH_0.3_G004849	Muscle-related coiled-coil protein	control vs T8day	18.6

FC = fold change in gene expression; positive values represent upregulation, and negative values indicate downregulation, in the treatment that involved longer exposures to 17 α MT

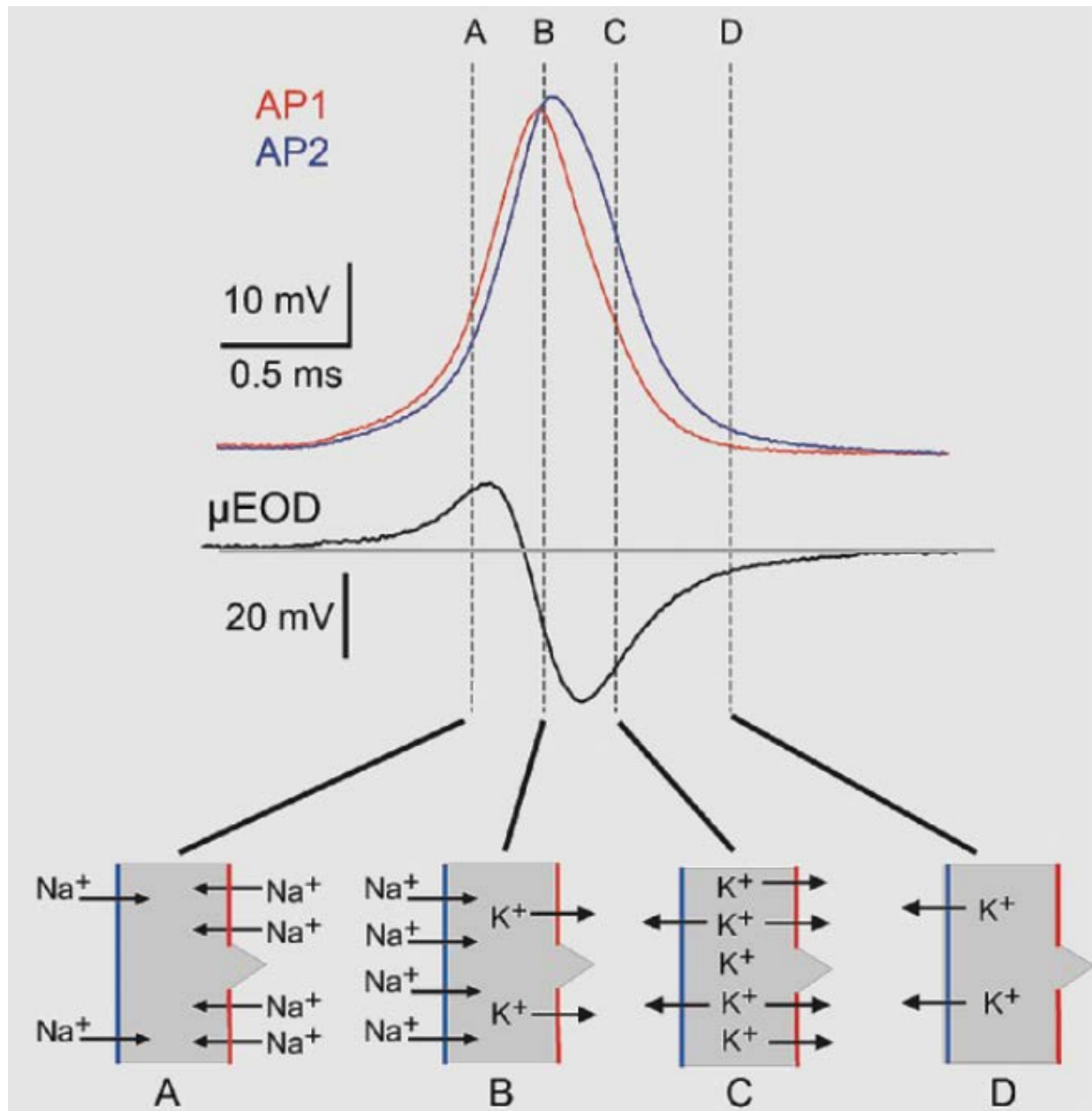


Figure 2.1. Generalized model of a biphasic EOD and its underlying APs. Top: APs of the posterior (AP1, red) and anterior (AP2, blue) faces of the electrocyte, and the resulting EOD (black, calculated as AP1 – AP2). Bottom: Cartoon electrocytes (posterior face red, anterior face blue) illustrating the Na⁺ and K⁺ ionic currents at key moments in the EOD. Adapted from (Stoddard and Markham 2008).

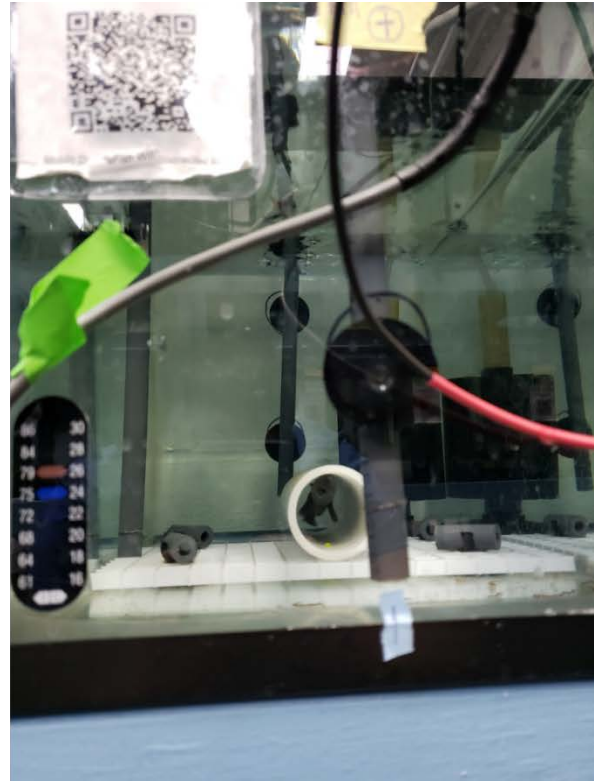
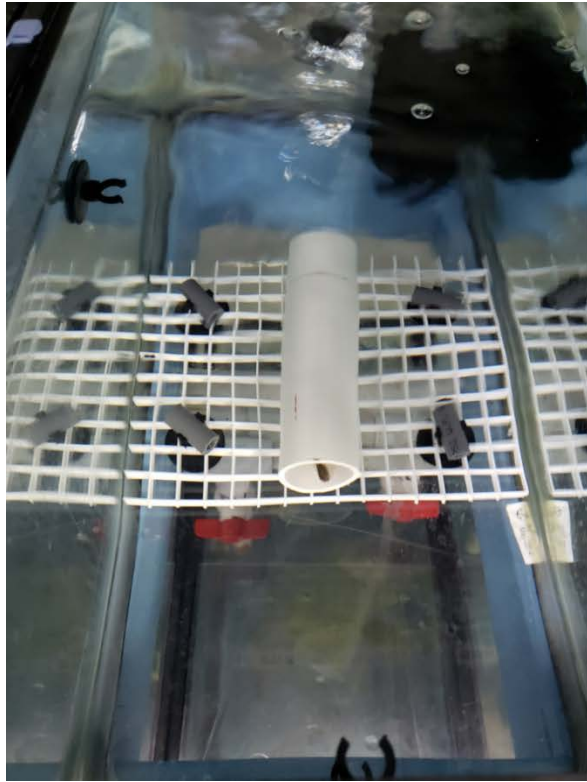


Figure 2.2. Photographs of the experimental fish tanks' configuration. A top view (left) shows the housing arrangement attached to the middle of the tank's bottom, a frontal view (right) illustrates electrode positioning and attachment during recording sessions.

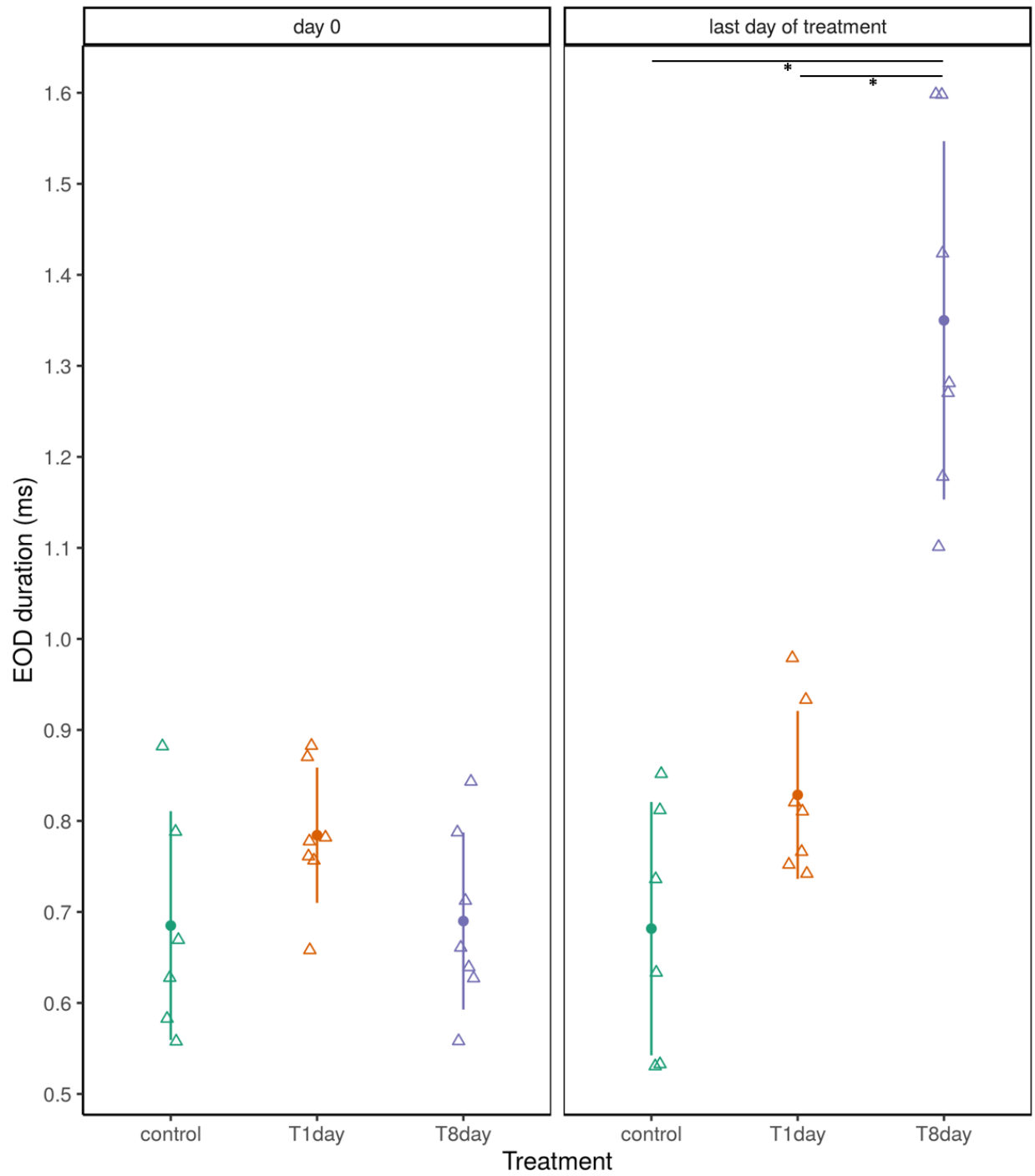


Figure 2.3. EOD duration at the onset (left) and at the end (right) of the experiment. The circle and vertical lines represent the mean EOD duration \pm standard deviation per treatment, open triangles are per fish measurements. A small horizontal jitter was added to better visualize overlapping observations. Each asterisk represents significant differences ($p < 0.001$) between the treatments connected by its overlying horizontal bar.

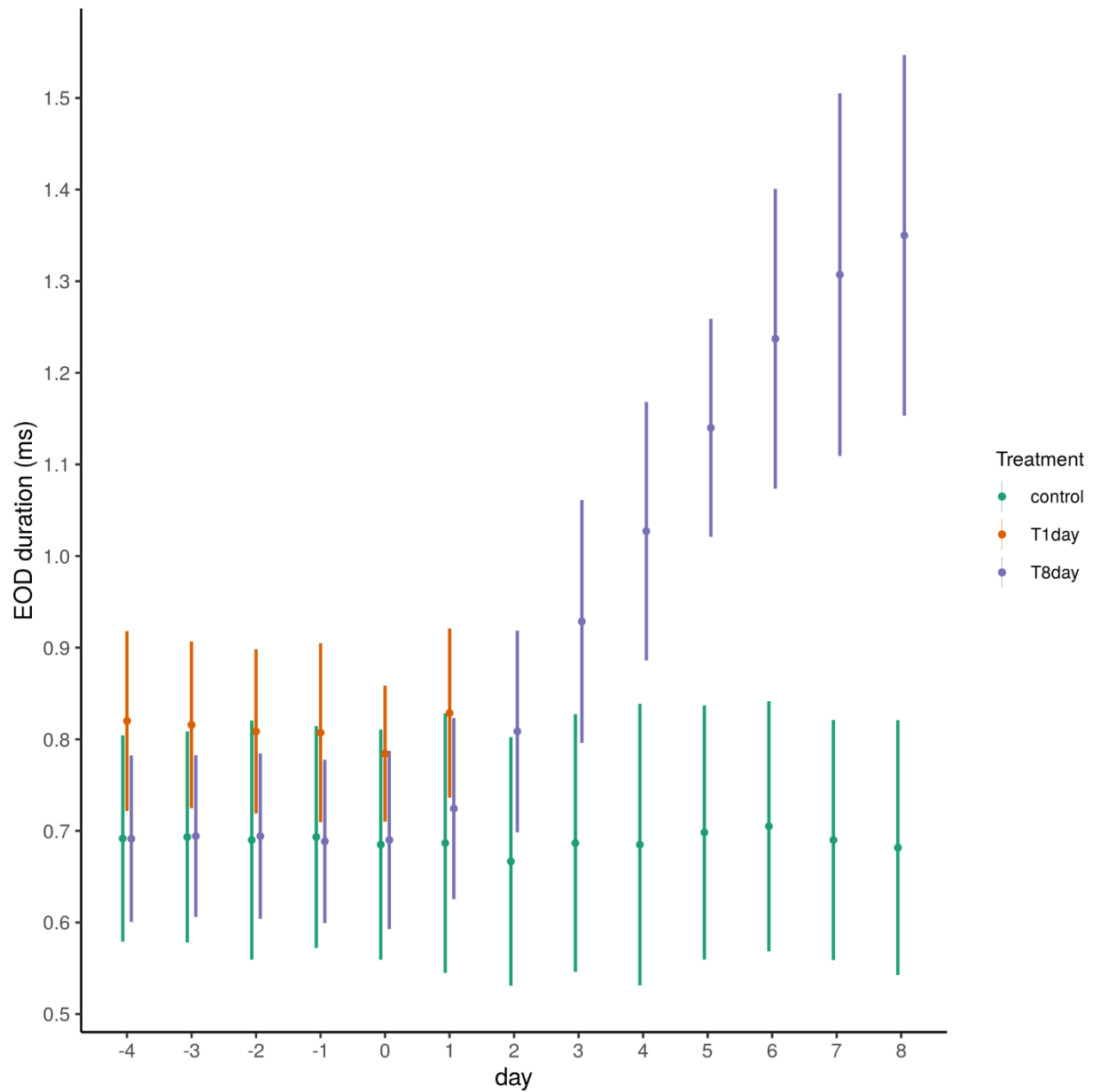


Figure 2.4. EOD duration per treatment throughout the experiment. Each colored circle and its vertical lines represent the mean EOD duration \pm standard deviation per treatment and day. A small horizontal jitter was added to better visualize overlapping values. Days -4 to 0 are part of the acclimation period, treatment-specific manipulations were performed on Day 0 after taking EOD recordings.

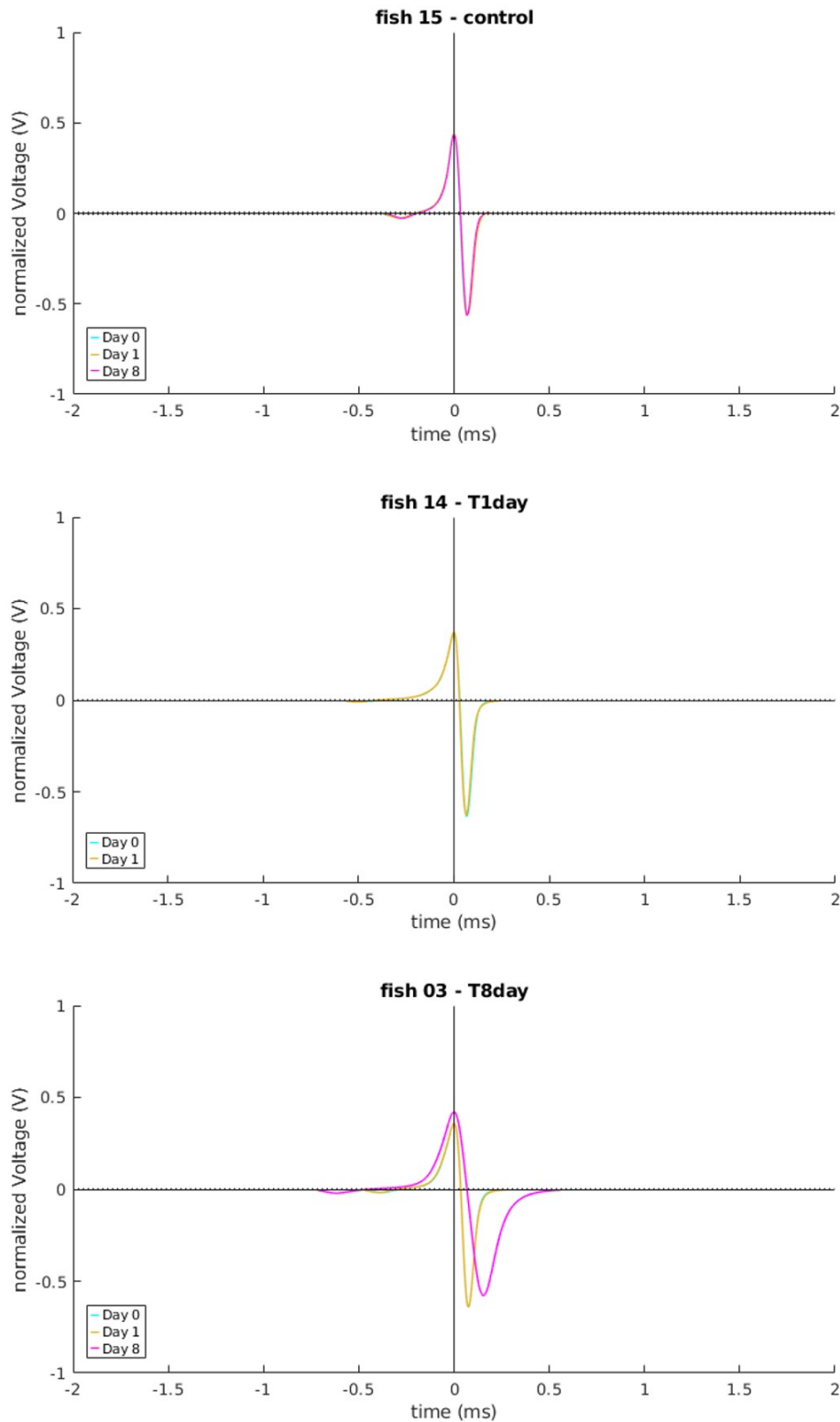


Figure 2.5. Change in EOD duration in one representative fish per treatment. Each graph shows the traces of the averaged EODs from individual fish recordings taken on Days 0, 1, and 8 (not applicable for treatment T1day). When EOD duration changes are very small, traces from more recent days overlap older traces.

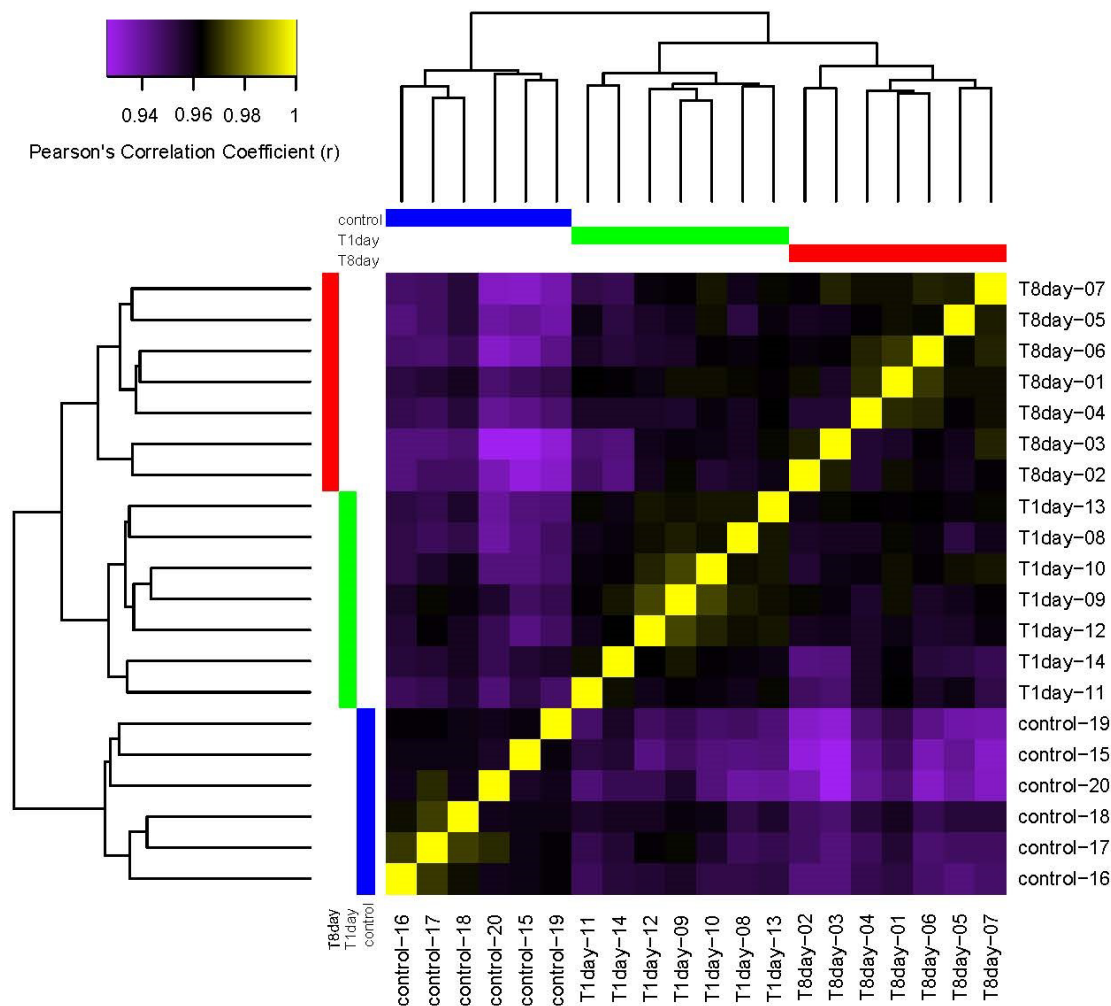


Figure 2.6. Heatmap of sample by sample correlations in gene expression, and the inferred relationships among treatments from these expression correlation values.

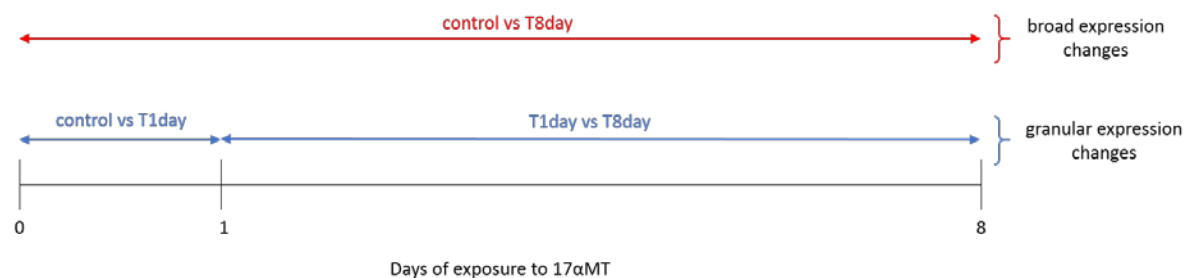


Figure 2.7. Cartoon depiction of how gene expression changes detected in the three pairwise treatment comparisons relate to each other in terms of time of exposure to 17αMT.

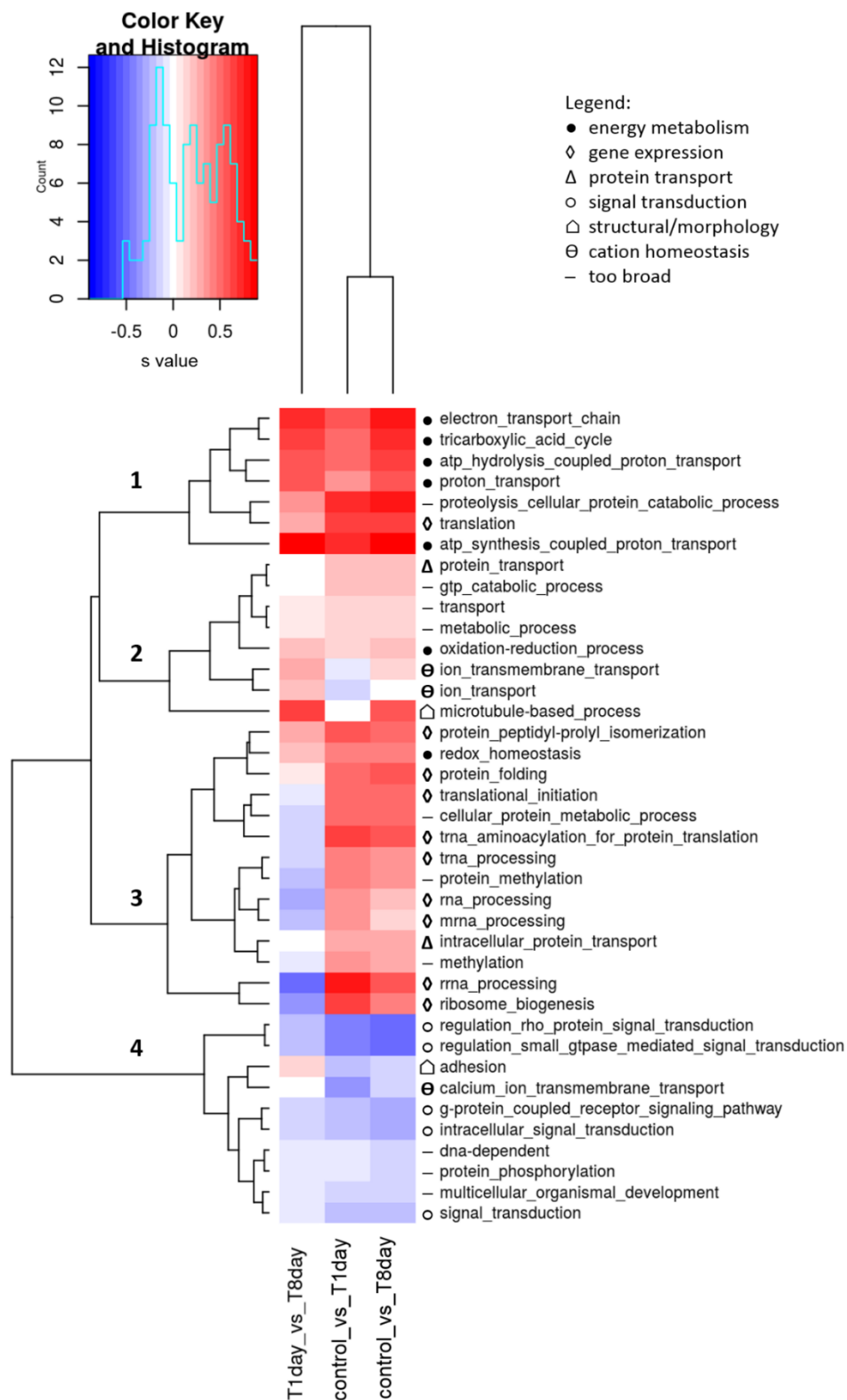


Figure 2.8. Heatmap of match enrichment scores (s) for all ($n = 39$) significantly enriched gene sets (higher dimensional enrichment score > 0.1 , FDR $p < 0.01$) in the three pairwise treatment contrasts studied. Gene sets are further classified into general categories (legend).

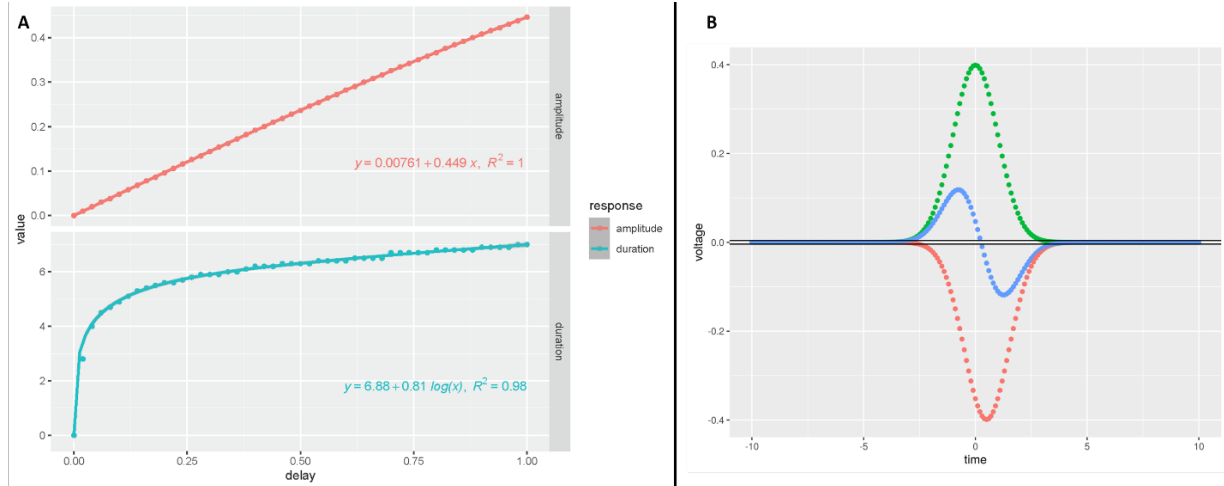


Figure 2.9. Simulated effects on EOD amplitude and duration of delayed depolarizations of the anterior electrocyte face in relation to the posterior face. The AP of each face is simulated as a normal distribution of $sd = 1$. Threshold for EOD detection = 0.004. A) EOD amplitude (top) and duration (bottom) across delays ranging from 0 to 1. Units and values are irrelevant. B) a representative EOD plot with delay = 0.50. Green = AP of posterior face, red = AP of anterior face, blue = resulting EOD, black horizontal lines = EOD detection threshold.

REFERENCES

REFERENCES

- Arnegard, M. E., S. M. Bogdanowicz, and C. D. Hopkins. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. *Evolution* 59:324–343.
- Arnegard, M. E., and C. D. Hopkins. 2003. Electric signal variation among seven blunt-snouted *Brienomyrus* species (Teleostei: Mormyridae) from a riverine species flock in Gabon, Central Africa. *Environ. Biol. Fishes* 67:321–339.
- Arnegard, M. E., B. S. Jackson, and C. D. Hopkins. 2006. Time-domain signal divergence and discrimination without receptor modification in sympatric morphs of electric fishes. *J. Exp. Biol.* 209:2182–2198.
- Arnegard, M. E., P. B. McIntyre, L. J. Harmon, M. L. Zelditch, W. G. R. Crampton, J. K. Davis, J. P. Sullivan, S. Lavoué, and C. D. Hopkins. 2010a. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *Am. Nat.* 176:335–356.
- Arnegard, M. E., D. J. Zwickl, Y. Lu, and H. H. Zakon. 2010b. Old gene duplication facilitates origin and diversification of an innovative communication system--twice. *Proc. Natl. Acad. Sci. U.S.A.* 107:22172–22177.
- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29.
- Bass, A. H. 1986. Species differences in electric organs of mormyrids: Substrates for species-typical electric organ discharge waveforms. *J. Comp. Neurol.* 244:313–330.
- Bass, A. H., J.-P. Denizot, and M. A. Marchaterre. 1986. Ultrastructural features and hormone-dependent sex differences of mormyrid electric organs. *J. Comp. Neurol.* 254:511–528.
- Bass, A. H., and C. D. Hopkins. 1985. Hormonal control of sex differences in the electric organ discharge (EOD) of mormyrid fishes. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 156:587–604.
- Bass, A. H., and C. D. Hopkins. 1983. Hormonal control of sexual differentiation: changes in electric organ discharge waveform. *Science* 220:971–974.
- Bass, A. H., and S. F. Volman. 1987. From behavior to membranes: testosterone-induced changes in action potential duration in electric organs. *Proc. Natl. Acad. Sci. U.S.A.* 84:9295–9298.
- Bennett, M. V. L. 1971. Electric Organs. Pp. 347–491 in W. S. Hoar and D. J. Randall, eds. *Fish Physiology*. Academic Press, London.
- Bennett, M. V. L., and H. Grundfest. 1961. Studies on the morphology and electrophysiology of electric organs. III. Electrophysiology of electric organs in mormyrids. Pp. 113–35 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, Amsterdam.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Carlson, B. A. 2002. Electric signaling behavior and the mechanisms of electric organ discharge production in mormyrid fish. *J. Physiol. Paris* 96:405–419. Elsevier.

Carlson, B. A., S. M. Hasan, M. Hollmann, D. B. Miller, L. J. Harmon, and M. E. Arnegard. 2011. Brain evolution triggers increased diversification of electric fishes. *Science* 332:583–586.

Carlson, B. A., C. D. Hopkins, and P. Thomas. 2000. Androgen correlates of socially induced changes in the electric organ discharge waveform of a mormyrid fish. *Horm. Behav.* 38:177–186.

Dunlap, K. D., A. C. Silva, G. T. Smith, and H. H. Zakon. 2017. Weakly Electric Fish: Behavior, Neurobiology, and Neuroendocrinology. Pp. 69–98 *in* D. W. Pfaff and M. Joëls, eds. *Hormones, Brain and Behavior*. Academic Press, Oxford.

Ferrari, M. B., M. L. McAnelly, and H. H. Zakon. 1995. Individual variation in and androgen-modulation of the sodium current in electric organ. *J. Neurosci.* 15:4023–4032.

Ferrari, M. B., and H. H. Zakon. 1993. Conductances contributing to the action-potential of *Sternopygus electrocytes*. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 173:281–292. Springer.

Few, W. P., and H. H. Zakon. 2007. Sex differences in and hormonal regulation of Kv1 potassium channel gene expression in the electric organ: Molecular control of a social signal. *Dev. Neurobiol.* 67:535–549. John Wiley & Sons.

Freedman, E. G., J. Olyarchuk, M. A. Marchaterre, and A. H. Bass. 1989. A temporal analysis of testosterone-induced changes in electric organs and electric organ discharges of mormyrid fishes. *J. Neurobiol.* 20:619–634.

Gallant, J. R., M. E. Arnegard, J. P. Sullivan, B. A. Carlson, and C. D. Hopkins. 2011. Signal variation and its morphological correlates in *Paramormyrops kingsleyae* provide insight into the evolution of electrogenic signal diversity in mormyrid electric fish. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 197:799–817.

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

Gallant, J. R., L. L. Traeger, J. D. Volkening, H. Moffett, P.-H. Chen, C. D. Novina, G. N. Phillips, R. Anand, G. B. Wells, M. Pinch, R. Guth, G. A. Unguez, J. S. Albert, H. H. Zakon, M. P. Samanta, and M. R. Sussman. 2014. Genomic basis for the convergent evolution of electric organs. *Science* 344:1522–1525.

García-Campos, M. A., J. Espinal-Enríquez, and E. Hernández-Lemus. 2015. Pathway analysis: State of the art. *Front. Physiol.* 6:1–16.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from

RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–52. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–512. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

Hall, A. 1998. Rho GTPases and the actin cytoskeleton. *Science* 279:509–514.

Herfeld, S., and P. Moller. 1998. Effects of 17 α -methyltestosterone on sexually dimorphic characters in the weakly discharging electric fish, *Brienomyrus niger* (Gunther, 1866) (Mormyridae): Electric organ discharge, ventral body wall indentation, and anal-fin ray bone expansion. *Horm. Behav.* 34:303–319.

Holt, C., and M. Yandell. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12.

Hopkins, C. D. 1986. Behavior of Mormyridae. Pp. 527–576 in T. H. Bullock and W. Heiligenberg, eds. *Electroreception*. Wiley, New York.

Hopkins, C. D. 1999a. Design features for electric communication. *J. Exp. Biol.* 202:1217–1228. *Co Biol.*

Hopkins, C. D. 1981. On the diversity of electric signals in a community of mormyrid electric fish in West Africa. *Am. Zool.* 21:211–222. Oxford University Press.

Hopkins, C. D. 1999b. Signal evolution in electric communication. Pp. 461–491 in M. Hauser and M. Konishi, eds. *The design of animal communication*. MIT Press, Cambridge, MA.

Hopkins, C. D., and A. H. Bass. 1981. Temporal coding of species recognition signals in an electric fish. *Science* 212:85–87.

Kaspi, A., and M. Ziemann. 2020. Mitch: Multi-contrast pathway enrichment for multi-omics and single-cell profiling data. *BMC Genomics* 21:1–17. BMC Genomics.

Kornienko, Y., R. Tiedemann, M. Vater, and F. Kirschbaum. 2021. Ontogeny of the electric organ discharge and of the papillae of the electrocytes in the weakly electric fish *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *J. Comp. Neurol.* 529:1052–1065.

Kramer, B. 1974. Electric organ discharge interaction during interspecific agonistic behaviour in freely swimming mormyrid fish. *J. Comp. Physiol.* 93:203–235. Springer-Verlag.

Lamanna, F., F. Kirschbaum, I. Waurick, C. Dieterich, and R. Tiedemann. 2015. Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). *BMC Genomics* 16:668. BMC Genomics.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359. Nature Publishing Group.

- Lavoué, S., M. E. Arnegard, J. P. Sullivan, and C. D. Hopkins. 2008. Petrocephalus of Odzala offer insights into evolutionary patterns of signal diversification in the Mormyridae, a family of weakly electrogenic fishes from Africa. *J. Physiol. Paris* 102:322–339. Elsevier.
- Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. BioMed Central.
- Lien, C. C., and P. Jonas. 2003. Kv3 potassium conductance is necessary and kinetically optimized for high-frequency action potential generation in hippocampal interneurons. *J. Neurosci.* 23:2058–2068.
- Lissmann, H. W., and K. E. Machin. 1958. The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J. Exp. Biol.* 35:451–486.
- Liu, H., M. M. Wu, and H. H. Zakon. 2007. Individual variation and hormonal modulation of a sodium channel β subunit in the electric organ correlate with variation in a social signal. *Dev. Neurobiol.* 67:1289–1304. John Wiley & Sons.
- Losilla, M., D. M. Luecke, and J. R. Gallant. 2020. The transcriptional correlates of divergent electric organ discharges in *Paramormyrops* electric fish. *BMC Evol. Biol.* 20:6.
- Markham, M. R. 2013. Electrocyte physiology: 50 years later. *J. Exp. Biol.* 216:2451–2458.
- Markham, M. R., and H. H. Zakon. 2014. Ionic Mechanisms of Microsecond-Scale Spike Timing in Single Cells. *J. Neurosci.* 34:6668–6678.
- Mederos y Schnitzler, M., S. Rinné, L. Skrobek, V. Renigunta, G. Schlichthörl, C. Derst, T. Gudermann, J. Daut, and R. Preisig-Müller. 2009. Mutation of histidine 105 in the T1 domain of the potassium channel Kv2.1 disrupts heteromerization with Kv6.3 and Kv6.4. *J. Biol. Chem.* 284:4695–4704.
- Mills, A., and H. H. Zakon. 1987. Coordination of EOD frequency and pulse duration in a weakly electric wave fish: the influence of androgens. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 161:417–430. Springer-Verlag.
- Möhres, F. P. 1957. Elektrische Entladungen im Dienste der Revierabgrenzung bei Fischen. *Naturwissenschaften* 44:431–432. Springer-Verlag.
- Nagel, R., F. Kirschbaum, and R. Tiedemann. 2017. Electric organ discharge diversification in mormyrid weakly electric fish is associated with differential expression of voltage-gated ion channel genes. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 203:183–195. Springer Berlin Heidelberg.
- Nguyen, L., V. Mamonekene, M. Vater, P. Bartsch, R. Tiedemann, and F. Kirschbaum. 2020. Ontogeny of electric organ and electric organ discharge in *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 206:453–466. Springer Berlin Heidelberg.
- Paul, C., V. Mamonekene, M. Vater, P. G. D. Feulner, J. Engelmann, R. Tiedemann, and F. Kirschbaum. 2015. Comparative histology of the adult electric organ among four species of the genus *Campylomormyrus* (Teleostei: Mormyridae). *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 201:357–374.

- Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4:1–8.
- Robinson, M. D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.
- Rudy, B., and C. J. McBain. 2001. Kv3 channels: Voltage-gated K⁺ channels designed for high-frequency repetitive firing. *Trends Neurosci.* 24:517–526.
- Salazar, V. L., and P. K. Stoddard. 2008. Sex differences in energetic costs explain sexual dimorphism in the circadian rhythm modulation of the electrocommunication signal of the gymnotiform fish *Brachyhypopomus pinnicaudatus*. *J. Exp. Biol.* 211:1012–1020.
- Schwartz, I. R., G. D. Pappas, and M. V. L. Bennett. 1975. The fine structure of electrocytes in weakly electric teleosts. *J. Neurocytol.* 4:87–114.
- Shenkel, S., and F. J. Sigworth. 1991. Patch recordings from the electrocytes *Electrophorus electricus*: Na currents and PNa/PK variability. *J. Gen. Physiol.* 97:1013–1041. Rockefeller Univ Press.
- Stoddard, P. K., and M. R. Markham. 2008. Signal Cloaking by Electric Fish. *Bioscience* 58:415.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2002. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution* 56:597–616. Society for the Study of Evolution.
- Swapna, I., A. Ghezzi, J. M. York, M. R. Markham, D. B. Halling, Y. Lu, J. R. Gallant, and H. H. Zakon. 2018. Electrostatic Tuning of a Potassium Channel in Electric Fish. *Curr. Biol.* 28:2094–2102.e5. Cell Press.
- Szabo, T. 1961. Les Organes Electriques des Mormyrides. Pp. 20–24 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, New York.
- The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47:D506–D515.
- von der Emde, G., M. Amey, J. Engelmann, S. Fetz, C. Folde, M. Hollmann, M. Metzen, and R. Pusch. 2008. Active electrolocation in *Gnathonemus petersii*: Behaviour, sensory performance, and receptor systems. *J. Physiol. Paris* 102:279–290. Elsevier Ltd.
- Zakon, H. H., Y. Lu, D. J. Zwickl, and D. M. Hillis. 2006. Sodium channel genes and the evolution of diversity in communication signals of electric fishes: Convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 103:3675–3680. Section of Neurobiology, University of Texas, Austin, TX 78712, USA. h.zakon@mail.utexas.edu.
- Zakon, H. H., A. C. Mills, and M. B. Ferrari. 1991. Androgen-dependent modulation of the electrosensory and electromotor systems of a weakly electric fish. *Semin. Neurosci.* 3:449–457.

Introduction

A major theme in evolutionary biology is the discernment of the evolutionary forces and molecular mechanisms that underlie phenotypic variation, and after half a century of research and discussion, the roles of positive selection and neutral evolution remain contentious (Kimura 1968; King and Jukes 1969; Kreitman 1996; Ohta 1996; Nei et al. 2010; Kern and Hahn 2018; Jensen et al. 2019). Furthermore, elucidating the contributions of specific genes to phenotypic variation and divergence of animal communication signals is complicated by intrinsic features of behaviors. They are context-dependent, often difficult to interpret, involve multiple traits, are frequently under polygenic control (Hoy et al. 1977; Shaw and Lesnick 2009; Wiley et al. 2012; Albertson et al. 2014; Ding et al. 2014), and require coordinated functioning between sensory and motor systems (Zakon et al. 2008; Straka et al. 2018). These signals must diverge rapidly during explosive species diversifications (Mendelson and Shaw 2005; Arnegard et al. 2006; Mason et al. 2017).

African weakly electric fish (Mormyridae) are among the most rapidly speciating clades of ray-finned fishes (Carlson et al. 2011; Rabosky et al. 2013). This is fueled in part by the ongoing explosive diversification of the genus *Paramormyrops* (Sullivan et al. 2002, 2004) in West-Central Africa, where no less than 20 species (Lavoué et al. 2008) have evolved, most within the last 500 000 years (Sullivan et al. 2002). Mormyrid electric organ discharges (EODs) are a central constituent of their electrolocation (Lissmann and Machin 1958; von der Emde et al. 2008) and intraspecific communication (Möhres 1957; Kramer 1974) behaviors. Often EODs are the most reliable method for discriminating between species (Hopkins 1981; Bass 1986; Alves-Gomes and Hopkins 1997; Sullivan et al. 2000; Arnegard et al. 2005), a pattern acutely evident in *Paramormyrops* (Picq et al. 2020), in which EOD waveform evolution greatly outpaces that of morphology, size, and trophic ecology (Arnegard et al. 2010). EOD waveform complexity refers to the number of phases present in the EOD, and it is a product of the ultrastructural configuration of the electrocytes, the cells that make up the electric organ (for details see (Bennett and

Grundfest 1961; Szabo 1961; Bennett 1971; Hopkins 1999; Gallant et al. 2011; Carlson and Gallant 2013)). We recently identified gene expression correlates for variable EOD features by comparing 11 specimens from five *Paramormyrops* species (Losilla et al. 2020). For biphasic and triphasic phenotypes of EOD waveform complexity, the most remarkable gene expression correlates were two copies of the gene *epdl2* (ependymin-like 2, named after its zebrafish ortholog), a member of the ependymin-related family (EPDR). Here, we elucidate the molecular evolutionary history of *epdl2* in Mormyridae, with emphasis on *Paramormyrops*.

EPDR proteins are secretory, calcium-binding glycoproteins. They form homodimers (Shashoua 1985; Hoffmann 1994) but can construct higher molecular weight aggregates (Shashoua et al. 1990), change their structural conformation according to calcium concentration (Ganss and Hoffmann 2009), and participate in cell-cell and cell-extracellular matrix interactions (Schwarz et al. 1993; Hoffmann 1994; Pradel et al. 1999). Despite low amino acid conservation, they share similar hydropathy profiles, N-glycosylation sites, and conserved cysteine residues (Ortí and Meyer 1996). Presumably, these biochemical and structural properties allow EPDRs to participate in an extraordinarily diverse range of functions, often lineage specific. They have been implicated in memory formation and neuroplasticity in fish (Shashoua 1985; Schmidt et al. 1995; Pradel et al. 1999), shell patterning and pigmentation in gastropods (Jackson et al. 2006), intestinal regeneration in sea cucumbers (Zheng et al. 2006), collagen contractibility of human fibroblasts (Staats et al. 2016), and conspecific communication in crown-of-thorns starfish (Hall et al. 2017). A systematic survey of eukaryotic sequences on EPDR proteins identified regions with little sequence similarity, but also detected a conserved signal peptide, N-glycosylation sites, several conserved residues including four key cysteines expected to form two disulfide bonds, a conserved protein length (median size 213 amino acids); and predicted conserved 3D structures that include antiparallel β sheets that fold into a pocket with suggested ligand-binding functions (McDougall et al. 2018). Recent studies independently determined the crystal structure of

EPDR1 in human (Wei et al. 2019), and human, mouse and the western clawed frog (*Xenopus tropicalis*) (Park et al. 2019). They confirmed that this protein forms disulfide bonds, associates into homodimers, and contains a ligand-binding hydrophobic pocket central to the protein's function; which is likely to involve lipid transport and/or metabolism (Park et al. 2019; Wei et al. 2019). Although no EPDRs have been identified outside eukaryotes, the structure of EPDR1's hydrophobic pocket presents unambiguous topological similarities with the functions-rich bacterial proteins of the LolA superfamily, thus implying an extensive distribution of this fold throughout cellular life (Park et al. 2019; Wei et al. 2019).

Notably, the *Paramormyrops kingsleyae* genome suggests the presence of three *epdl2* paralogs, whereas only one copy is present in the osteoglossiform *Scleropages formosus* and in the mormyrid *Brienomyrus brachyistius* (unpublished genome, v0.3), which is relatively basal compared to *Paramormyrops* (Alves-Gomes and Hopkins 1997; Sullivan et al. 2000; Lavoué et al. 2003). This pattern, along with the aforementioned gene expression correlates with EOD complexity, beg the question of if *epdl2* copy number or sequence evolution have been a factor in electric signal diversification in *Paramormyrops* or Mormyridae. Gene duplications can be the source of new or specialized functions through functional divergence of the duplicated copies (Taylor and Raes 2004; Conant and Wolfe 2008; Chapal et al. 2019) and this process can contribute to the emergence of species-specific characteristics and taxon diversification (Zhang 2003). The EPDR family has undergone extensive independent tandem duplications and divergence within several metazoan groups (McDougall et al. 2018), so much that EPDR paralogs have been proposed as suitable targets to experimentally test gene subfunctionalization (Suárez-Castillo and García-Arrarás 2007).

In this study, we detail a molecular evolutionary analysis of *epdl2* in African weakly electric fish. Since the genome annotations of both the published (Gallant et al. 2017) and unpublished (Nanopore assembly) *P. kingsleyae* genome assemblies contain three incomplete *epdl2* copies, we start by confirming the existence of three *epdl2* paralogs in *P. kingsleyae*. Then we sequence and identify *epdl2*

gene copies across Mormyridae, with emphasis on *Paramormyrops*. We elucidate paralog diversity, infer duplications history, and test for signatures of natural selection on branches and sites. Finally, we model the 3D structure of a representative *epdl2* protein and propose how the sites deemed under the strongest positive selection in paralogs may affect protein function.

Methods

epdl2 sequences from Osteoglossiformes genomes

We gathered *epdl2* gene sequences from genome assemblies of *S. formosus*, *Gymnarchus niloticus* (NCBI, Table 3.1) and *B. brachyistius* (unpublished). We identified the *epdl2* genes with BLAST searches against each genome, using *epdl2* genes (blastn) and proteins (tblastn) as queries. For *B. brachyistius*, the *epdl2* queries were the incomplete *epdl2* gene models from the *P. kingsleyae* Nanopore assembly; and once identified, the *B. brachyistius epdl2* gene served as query for the *G. niloticus* and *S. formosus* searches.

Confirm *epdl2* triplication in *Paramormyrops kingsleyae*

epdl2 paralogs in the *P. kingsleyae* Nanopore assembly

We used the sequences of the *epdl2* paralogs, along with flanking genes *otu1b* and *nrg2*, from the NCBI-annotated (Release 100) *P. kingsleyae* genome (Gallant et al. 2017), to find the *epdl2* genes in the *Paramormyrops kingsleyae* Nanopore assembly. All queries mapped to one gene region, where the NCBI annotation suffered from insufficient scaffolding, and the Nanopore assembly contained frame shifts caused by small indels in coding homopolymers. These structural problems impeded complete and accurate *epdl2* annotations in each assembly. As such, we decided to amplify and Sanger-sequence each potential paralog to confirm the presence of three *epdl2* paralogs and to deduce each paralog's gene structure and coding sequence.

PCR amplification and Sanger sequencing of *epdl2* paralogs in *P. kingsleyae*

We designed primers to amplify each *P. kingsleyae epdl2* paralog (Table 3.2). Their binding sites are 500-1500 bp up or downstream of the start/stop codon of each paralog, where primer binding was

predicted to amplify a unique PCR product. The targeted areas encompass the 5' and 3' UTRs, which are the ideal targets to design primers to amplify all *epd12* paralogs across taxa, at a later step.

For PCR amplifications we used Q5 High-Fidelity DNA Polymerase (NEB, cat. M0491S) and the same *P. kingsleyae* DNA used to sequence the Nanopore assembly. We carried out PCRs in 50 µl reactions, with the reagents and conditions detailed in Table 3.3 and Table 3.4. PCR products were cleaned (Monarch PCR & DNA Cleanup Kit, NEB cat. T1030S) and Sanger-sequenced. For the latter, in addition to the PCR primers, we designed sequencing primers (Table 3.2) that bind to regions conserved across all three putative paralogs and were spaced by < ~700bp. Thus, each amplified *epd12* paralog was sequenced with six or seven Sanger-sequencing reactions. Finally, we used the obtained paralog-specific sequences and the *epd12* annotations from *B. brachyistius* to produce gene models for each *epd12* paralog in *P. kingsleyae*.

epd12 sequences across Mormyridae

Primer design

We designed primers to target all *epd12* paralogs in a given mormyrid species (Table 3.2). As primer-binding regions, we chose the 5' and 3' UTRs, close to the start and stop codons. Primer design was guided by the Sanger-sequenced *P. kingsleyae* *epd12* paralogs, our *Brienomyrus brachyistius* *epd12* sequence, and whole genome sequencing scaffolds from species across Mormyridae, kindly provided by Rose Peterson and Stacy Pirro (Iridian Genomes) (now deposited at NCBI, Table 3.5). We assembled the target UTRs from whole genome sequencing scaffolds i) returned from BLAST searches (tblastn) of *P. kingsleyae* and *B. brachyistius* *Epd12* proteins against the scaffolds, or ii) that mapped to the *P. kingsleyae* or *B. brachyistius* *epd12* genomic region.

DNA extraction, PCR amplification and *epd12* sequencing

Individual mormyrid specimens were collected during field trips to Gabon, West Central Africa in 2019. Specimens were collected using hand nets combined with electric fish detectors. EODs of each specimen were originally recorded within 24 hours of capture in 1- to 5-liter plastic boxes filled with

water from the collection site. Signals were recorded with bipolar chloride-coated silver wire electrodes and amplified (bandwidth =0.0001-50 kHz) with a differential bioamplifier (CWE, Inc), and digitized at a 1 MHz sampling rate, with head-positivity plotted upward using a or a USB-powered AC-DC Converter (National Instruments). All EOD recordings were made at a vertical resolution of 16 bits per sample. For each specimen, 10 individual EOD waveforms were recorded. After recording their EODs, individual specimens were euthanized with an overdose of MS-222, and one or more paired fins from each were clipped and preserved in 95% ethanol. Each specimen was given a unique specimen identification tag, and fixed free-floating in 10% formalin (phosphate-buffered; pH 7.2) for at least 2 weeks. Specimens were then transferred to 70% ethanol and deposited in the Michigan State University Museum of Vertebrates. All methods conform to protocols approved by Michigan State University's Office of Environmental Health & Safety, and Institutional Animal Care & Use Committee

We extracted DNA from the ethanol-preserved fin clips, from one individual of each of our studied species (Table 3.1), with the DNeasy Blood & Tissue Kit (Qiagen, cat. 69504). Then we targeted all *epd12* paralogs for PCR amplification in 25 µl reactions, with primer- and species-dependent reagents (Table 3.6) and conditions (Table 3.7). The resulting amplicons were multiplexed and sequenced on an Oxford Nanopore Technologies (ONT) MinION device with a R9.4.1 flow cell (Nanopore sequencing). The multiplexed sequencing library was prepared with ONT kits: Native Barcoding Expansion 1-12 (cat. EXP-NBD104) & 13-24 (cat. EXP-NBD114), and Ligation Sequencing Kit (cat. SQK-LSK109); following manufacturer's directions.

Analysis of Nanopore reads

Read processing

Reads were base-called and quality-filtered ($Q > 7$) with the high accuracy model from Guppy 4.2.3+f90bd04 (ONT), and barcoded reads passing filtering were demultiplexed with Guppy 4.4.1+1c81d62. We performed all subsequent steps on these demultiplexed reads.

We modified the amplicon subcommand of seqkit 0.15.1 (Shen et al. 2016) to identify all amplicons in each read. Next, using a custom pipeline consisting of seqkit, Nanofilt 2.71 (De Coster et al. 2018) and custom code, we extracted the smallest amplicon greater than 1000 bp. Extracted amplicons were then subjected to one additional round of size (1200-2400 bp) and quality ($Q > 14$) filtering with NanoFilt. We generated quality control summaries and plots with Nanoplot 1.33.1 (De Coster et al. 2018).

Identification of *epd12* genes

The following bioinformatic procedure is summarized in Fig. 3.1. The filtered amplicons from each sample were clustered using cd-hit 4.8.1 (Li and Godzik 2006). A preliminary analysis of the data revealed a key parameter is cd-hit's sequence identity threshold (c , range 0-1). We noticed in *P. kingsleyae*, where the number of *epd12* copies were known through Sanger sequencing, that relatively low values of c lumped amplicons from different genes into the same cluster (underclustering), whereas relatively high values split amplicons from the same gene into different clusters (overclustering). To mitigate this issue, for each sample, we ran cd-hit with c values from 0.84 to 0.91 in 0.01 increments and selected the c value that produced the best clustering. We chose this range because the selected c value from every sample was higher than 0.84 and lower than 0.91. The c value that generated the best clustering was chosen as follows: we selected the c value that produced the largest number of supported clusters. If more than one c value met this criterion, we chose a c value that classified the largest number of amplicons into supported clusters. If more than one c value met this, we selected the largest c value. Once a c value was selected, we generated consensus sequences for each cluster and manually inspected them for overclustering, which in our dataset manifested as consensus sequences from different clusters with greater than 99.45% sequence identity. In these few cases, we reapplied the selection criteria to choose a new c value that produced fewer supported clusters than the discarded c value.

Following clustering with cd-hit, we obtained consensus sequences from each cluster that were supported by > 10% of the amplicons with Medaka 1.2.1 (smolecule module) (<https://github.com/nanoporetech/medaka>), and manually inspected the consensus sequences in Geneious 9.1.8 (Biomatters) by aligning amplicons to the *epdl2* sequences of *B. brachyistius* and *P. kingsleyae*. This allowed us to identify and remove non-specific amplicons. Finally, for each aligned consensus sequence, we manually annotated CDSs and edited homopolymer sequences in CDS4 to keep the sequence in frame. The code we used in this work can be found at <https://github.com/msuefishlab/>.

epdl2 gene tree and *epdl2* duplication history

We aligned all the mormyrid *epdl2* coding regions (= CDSs and introns) in Geneious with the Muscle method (Edgar 2004) followed by a manual inspection. We excluded the *epdl2* genes of *Gymnarchus niloticus* and *Scleropages formosus* because their very long introns 1 and 2 could not be aligned reliably. We used this alignment to infer a mormyrid *epdl2* gene tree with the maximum likelihood (ML) criterion through the online portal of PhyML 3.3 (Guindon et al. 2010) (<http://www.atgc-montpellier.fr/phyml>). The best-fitting nucleotide substitution model was determined by Smart Model Selection (SMS) 1.8.4 (Lefort et al. 2017) using the Akaike Information Criterion (AIC), and we estimated branch support with 10000 bootstrap replicates. The substitution model K80 + G best fitted the alignment of the *epdl2* coding regions, thus the *epdl2* gene tree was inferred with the model K80 + G + F.

Selection Tests

In order to test for signatures of natural selection, we extracted CDSs from all the sequences obtained from Nanopore sequencing, as well as additional osteoglossiform sequences (Table 3.1). Homologous codons were aligned by backtranslating aligned protein sequences on the online portal of TranslatorX (Abascal et al. 2010) (<http://translatorx.co.uk/>) using the Prank algorithm (Löytynoja and Goldman 2005). Then we manually refined alignment gaps and removed stop codons.

Next, we created an osteoglossiform *epdl2* gene tree by adding *G. niloticus* and *S. formosus* to the previously inferred mormyrid *epdl2* gene tree, and by making *Stomatorhinus* + *Pollimyrus* sister taxa. The evolutionary relationships of these four taxa within Osteoglossiformes are not disputed.

We ran six tests, using HyPhy (Kosakovsky Pond et al. 2005, 2020) via their online portal (Weaver et al. 2018) (<http://www.datamonkey.org>), except where indicated. First, we looked for evidence of recombination in the alignment of osteoglossiform *epdl2* CDSs with the method GARD 0.2 (Kosakovsky Pond et al. 2006). We ran this analysis three times, each with one of the three options available for site-to-site variation (none, general discrete, beta-gamma), and in every run the run mode was set to normal and rate classes was set to 4 (default). Second, to obtain a gene-wide overview, we used the BUSTED 3.1 algorithm (Murrell et al. 2015) with synonymous rate variation and all branches selected to test whether *epdl2* has experienced positive selection at any site or time (branch).

We then used a local installation of HyPhy 2.5.30 to run RELAX 4.0 (Wertheim et al. 2015) (synonymous rate variation allowed, omega rate classes = 2), to test whether the strength of selection on *epdl2* changed (relaxed or intensified) in the branches with *epdl2* duplications relative to the mormyrid branches without *epdl2* duplications. After this, we employed the branch-test aBSREL 2.2 (Smith et al. 2015) to identify branches that experienced positive selection in the time between the *epdl2* duplication events and the subsequent divergence into current *epdl2* paralogs. We limited this analysis to these branches because of three reasons: i) to conserve statistical power, since this test corrects for multiple testing, ii) the RELAX test suggested that the strength of selection has intensified in the duplicated lineages, and iii) to avoid the potential effects of the less-trustworthy terminal branch topology within each paralog clade. Since aBSREL is a branch-test, it is more sensitive to errors in the provided topology.

Finally, we attempted to identify evolutionarily interesting sites (codons) in *epdl2* genes. First, we used the site-test MEME 2.1.2 (Murrell et al. 2012) to detect which sites in *epdl2* have been

subjected to positive selection. Second, we used Contrast-FEL 0.5 (Kosakovsky Pond et al. 2021) to investigate which sites in *epdl2* may be evolving at different rates between the duplicated and the non-duplicated mormyrid lineages. The sites we are most interested in are those identified by both methods.

Structural predictions of Epd12

We chose the Epd12 amino acid sequence of *Paramormyrops szaboi* as a representative mormyrid Epd12 protein to explore structural features common in EPDR proteins. We searched for a signal peptide on the online portal of SignalP 5.0 (Almagro Armenteros et al. 2019) (www.cbs.dtu.dk/services/SignalP/), we identified N-Glycosylation sites on the web implementation of NetNGlyc 1.0 (Gupta and Brunak 2002) (www.cbs.dtu.dk/services/NetNGlyc/), and we predicted a 3D structure of the *epdl2* protein with its signal peptide removed using the I-TASSER server (Roy et al. 2010; Yang and Zhang 2015; Yang et al. 2015) (<https://zhanglab.dcm.b.med.umich.edu/I-TASSER/>). Finally, we annotated features and sites on the resulting PDB model with the web-based 3D structure viewer iCn3D (Wang et al. 2020) (<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>).

Results

Paramormyrops kingsleyae has three complete *epdl2* paralogs

We identified a ~40 Kbp region in chr24 in the *P. kingsleyae* Nanopore assembly with three *epdl2* copies in tandem (Fig. 3.2). According to the zebrafish, *Scleropages formosus*, and *Brienomyrus brachyistius* genomes, *epdl2* was originally located in the minus (-) strand.

Based on this genome sequence, we used PCR and Sanger sequencing to obtain amplicons of the expected sizes (~3200 and ~4000 bp, Fig. 3.3); and confirmed that there are three distinct *epdl2* paralogs in *P. kingsleyae*. Each of our three predicted gene models showcases the six expected CDSs, canonical splicing sites, start and stop codons, and a predicted protein size (213-214 amino acids) of standard EPDR length, with the four conserved cysteine residues, a highly conserved proline residue, and high sequence similarity to Epd12 proteins from other taxa, suggesting the all produce functional

proteins. We named these paralogs *epdl2.1*, *epdl2.2*, and *epdl2.3* (Fig. 3.2); their respective lengths from start to stop codon are: 1675, 1814, and 1809 bp.

epdl2 sequences and copy number across Mormyridae

We developed a Nanopore-based amplicon sequencing strategy to obtain full-length *epdl2* gene sequences across a variety of mormyrid species (Table 3.1), and used a custom bioinformatics pipeline to identify paralogs. As a validation of this approach, we included a *P. kingsleyae* (biphasic) sample in our analysis. Our pipeline found three *epdl2* paralogs in this sample, each shared 99.9% sequence identity from start to stop codon and the same predicted protein sequence to its Sanger sequenced counterpart. Every *epdl2* gene we identified with our experimental approach displays the hallmarks of a functional *epdl2* gene: six expected CDSs, canonical splicing sites, start and stop codons, and a predicted amino acid sequence of 213-215 amino acids, with the four conserved cysteine residues, a highly conserved proline residue, and high sequence similarity to *epdl2* proteins from other taxa.

When we applied this approach to other mormyrids, we identified a variable number of *epdl2* genes across various mormyrid species (Table 3.1), with a median size of 1815 bp (from start to stop codon). We were unable to obtain *epdl2* sequences in *Paramormyrops* SN2, *Paramormyrops* SN9, *Paramormyrops teugelsi*, and *Ivindomyrus marchei*. Using this approach, we were able to comprehensively survey *epdl2* sequences among Osteoglossiformes (*Scleropages formosus*), Gymnarchidae (*Gymnarchus niloticus*), and from the main branches of Mormyridae and *Paramormyrops*.

Paramormyrops underwent three *epdl2* duplications

Based on these sequences obtained, we generated a gene tree to infer the duplication history of *epdl2* (Fig. 3.4). The species relationships implied in the gene tree outside the *Marcusenius ntemensis* + *Paramormyrops* clade are largely congruent with previously estimated phylogenetic relationships, with the exception of a few branches with low bootstrap support (*Stomatorhinus ivindoensis*, *Pollimyrus* sp, and possibly *Brevimyrus niger*). *M. ntemensis* is widely regarded as a sister taxon to *Paramormyrops*

(Sullivan et al. 2002; Lavoué et al. 2003; Sullivan et al. 2004); however, our gene tree places this species within *Paramormyrops*.

For the majority of mormyrid taxa examined, there is only one identifiable *epdl2* gene present, in contrast to multiple paralogs found in *Marcusenius ntemensis* + *Paramormyrops* species (Table 3.1, Fig. 3.4). We detected only one *epdl2* gene in *Paramormyrops szaboi*, whereas all remaining *Paramormyrops* species examined, as well as *M. ntemensis*, either have multiple *epdl2* paralogs, or we identified only one copy that clearly belongs to a paralog clade. Paralogs homologous to *P. kingsleyae epdl2.1* branch earliest and share the unique feature of a 130-140 bp deletion in intron 2. A second group of *epdl2* paralogs bifurcates subsequently. This identifies a novel paralog, *epdl2.4*, that was not detected in the *Paramormyrops kingsleyae* genome. The remaining genes broadly sort into two groups homologous to *P. kingsleyae epdl2.2* and *epdl2.3*. We note that each of the four resulting *epdl2* paralog clades contains no more than one gene from any given sample. There are two branches in this gene tree that fall outside of these paralog clusters: *M. ntemensis epdl2.4* and *P. kingsleyae epdl2.2*. We assigned these putative identities based on inspection of their CDSs compared to unambiguously assigned paralogs, and on the shortest branch distances between these sequences and the paralog clades (Fig. 3.4). In the case of *P. kingsleyae epdl2.2*, the existence of *P. kingsleyae epdl2.3* indicates that the sequence in question does not belong to paralog *epdl2.3*.

Sequence divergence within duplicate *epdl2* genes is low, as reflected in their short branch lengths (Fig. 3.4). This divergence is lowest for paralogs *epdl2.1* and *epdl2.4*, and we detected both genes present in the same species in only two instances, *Paramormyrops curvifrons* and *M. ntemensis*. This may suggest that these two genes are allelic variants of the same paralog; however, we consider this unlikely because i) all *epdl2.1* sequences unambiguously and uniquely lack ~130 bp from intron 2, and ii) the predicted protein sequences of the two paralogs from *P. curvifrons*, and also those from *M.*

ntemensis, each share a pairwise sequence identity < 94%, likely too low to be alleles from the same gene.

Together, these results indicate that three rounds of *epd12* duplication have occurred within *Paramormyrops*. Given the broad distribution of species (Fig. 3.5) among the *epd12* paralogs, we propose that all *epd12* duplications occurred relatively quickly, in an ancestor common to all the *Paramormyrops* species sequenced except *P. szaboi*. We refer to this hypothetical ancestor as LABED, last ancestor before *epd12* duplications.

Selection Tests

EOD signal evolution within the more than 20 species of *Paramormyrops* is profound (Sullivan et al. 2002; Arnegard et al. 2010; Picq et al. 2020), and a previous study of differential gene expression among various signal types (Losilla et al. 2020) indicated a strong expression difference in *epd12.1* and *epd12.3* between biphasic and triphasic EODs. These results, together with the detection of as many as three rounds of *epd12* duplication in *Paramormyrops*, motivated us to search for evidence of selection on the *epd12* duplicates.

First, we examined for any evidence of recombination among *epd12* sequences that could lead to spurious results if not accounted for. We found no evidence of recombination in the alignment of osteoglossiform *epd12* CDSs after running three variants of the GARD method: in all cases the models with recombination showed no improvement over the no recombination null model, $\Delta c\text{-AIC} = 0$.

Next, we looked for evidence of gene-wide positive selection ($\omega > 1$) using the BUSTED method. We found that at least one site on at least one branch in the osteoglossiform *epd12* gene tree has experienced positive selection (LRT = 9.6, $p = 0.004$). The BUSTED test increases statistical power by pooling information across all sites and branches, at the expense of not pinpointing the sites/branches that drive the signal. Therefore, we deployed methods to parse patterns of selection in our data.

First, we compared ω values between duplicated and non-duplicated lineages (Fig. 3.6) using the RELAX method, and found strong indications that selection intensified in the duplicated lineages ($K = 7.6$,

LRT = 36.5, $p = 1.5 \times 10^{-9}$). Because this test compares ω branch values between user-defined branches and not against a value of one, it makes no claims about modes of selection. We then searched for signals of positive selection in the branches between the duplication events and the evolution of the extant paralog genes using the aBSREL test. Conservatively, we treated the difficult-to-group branches *Marcusenius ntemensis epdl2.4* and *Paramormyrops kingsleyae epdl2.2* as additional paralogs for the purpose of this test. Thus, we selected 11 branches for testing (ten were tested, aBSREL deleted one branch because it estimated it had zero length) and we found statistical support for positive selection on two of them (Fig. 3.7, branch A: LRT = 9.3, corrected $p = 0.03$; branch B: LRT = 8.7, corrected $p = 0.04$). These branches broadly represent the time between the second and third duplication, on the branch that eventually experienced the third duplication (*epdl2.2 + epdl2.3*).

Finally, we detected evolutionarily interesting sites in *epdl2*. We identified 21 *epdl2* codons that have experienced positive selection in one or more branches of the *epdl2* tree (MEME test, $p < 0.1$, Fig. 3.8). In addition, Contrast-FEL detected 43 sites evolving under different selective pressures between the duplicated and the non-duplicated mormyrid lineages (Fig. 3.9, $p < 0.1$), with all 43 sites presenting a higher ω value in the duplicated lineages. This method does not interrogate about modes of selection, instead it contrasts site-specific ω values between user-defined branches. From these results, we obtained a list of ten sites (Table 3.8) that i) have experienced positive selection in the *epdl2* gene tree (Fig. 3.8) and ii) have evolved at increased ω rates in duplicated vs non-duplicated mormyrid lineages (Fig. 3.9). For each positively selected site, MEME performs an exploratory attempt to identify the branches in which it has been under positive selection. It evaluates the Empirical Bayes Factor (EBF) for observing positive selection at each branch and flags a branch as potentially under positive selection if its EBF > 100 (Murrell et al. 2012) (Table 3.8). Given the experimental nature of this classification, we only consider it in relation to the results from the other tests we performed (Fig. 3.10). Finally, we summarize the amino acid substitutions observed at these ten sites in the *epdl2* paralogs (Table 3.9).

Structural predictions of Epd12

Because we were able to identify signatures of positive selection at particular *epd12* sites, we were motivated to link these sites to putative functional differences. To do so, we explored structural features using the Epd12 amino acid sequence of *Paramormyrops szaboi* as a representative mormyrid Epd12 protein. We identified a signal peptide (the first 17 amino acids) and two potential N-glycosylation sites. Also present are four conserved cysteines expected to form disulfide bonds and a highly conserved proline residue that may influence ligand binding ((McDougall et al. 2018)). We mapped these features and our ten sites of interest (Table 3.8) onto the primary structure of this reference protein (Fig. 3.11) and modeled its 3D structure with I-TASSER.

I-TASSER produces a full-length atomic model of the user-supplied sequence, based on the topology of the best-matching known structural template. Reassuringly, the chosen template was the crystal structure of *Xenopus tropicalis* Epd1 (Protein Data Bank (PDB) entry 6JL9), the only ependymin gene widely present in tetrapods (also known as MERP1). The resulting 3D model is supported with high confidence (C-score 0.59, TM-score 0.92 to PDB hit 6JL9, Fig. 3.12A-D). It is expected that Epd12 contains a hydrophobic pocket critical to its biological function, since this pocket is observed in Epd1 (Park et al. 2019; Wei et al. 2019) and modelled in all EPDR types (McDougall et al. 2018). In our model, this pocket is best visualized in Fig. 3.12D, outlined by antiparallel β sheets drawn in blue, cyan and yellow; covered with a “lid” depicted by the green β strand; and also covered in part by the C-terminal end (red, dark orange) which includes an α helix (red). Our 3D model prediction identified several potential ligand-binding residues and a few proposed active site residues that overlap with some of our sites of interest located in the pocket, although these predictions have low confidence.

We mapped the ten sites of interest (Table 3.8) on this 3D model: notably, these sites are located in the pocket lid (sites 105-106), its “lining” (sites 150, 153, 172, 176), and a long coil between two β strands (sites 125-7, 129) (Fig. 3.12E-F). According to our structural model, the side chains of these ten residues are oriented either inwards or outwards the hydrophobic pocket (Fig. 3.12E-F), which may

be indicative to their function: residues directed inwards may participate in ligand-binding (McDougall et al. 2018; Park et al. 2019; Wei et al. 2019) and/or calcium binding (Park et al. 2019); whereas the outward-facing residues belong to regions critical to Epdr1 dimerization in this protein's solved 3D structures (Park et al. 2019; Wei et al. 2019).

Discussion

The likely presence of duplicated *epdl2* paralogs in *Paramormyrops kingsleyae*, along with indications that some of the paralogs' expression in electric organs correlate with variation in key features of the electric signal ((Losilla et al. 2020)), motivated this molecular evolutionary study of *epdl2* in *Paramormyrops* and Mormyridae. We confirmed three extant, seemingly functional *epdl2* genes in *P. kingsleyae*. Then, we sequenced this gene across Mormyridae, identified four paralogs and found evidence for three tandem duplications in an early *Paramormyrops* ancestor. We detected signs of positive selection and of increased selection rates in branches that led to *epdl2* paralogs and in sites along the gene. Finally, we generated an *epdl2* 3D model and leveraged it to infer functional implications of amino acid substitutions at the sites esteemed under the strongest positive selection.

epdl2 amplicon sequencing

Our Sanger sequencing efforts confirm that there are three distinct, complete *epdl2* paralogs in tandem in *P. kingsleyae* (Fig. 3.2)—we provide here high-quality reference sequences for each gene, including intronic sequences and ample portions of the up- and downstream regions. Since we found no indication of additional paralogs in this region or elsewhere in the genome, we are confident that there are three *epdl2* paralogs in *P. kingsleyae*.

We were able to survey *epdl2* broadly across Mormyridae by developing a bespoke Nanopore-based amplicon sequencing approach and validated its sensitivity and accuracy with our Sanger sequenced *P. kingsleyae* genes. The combination of high-fidelity PCR, multiplexed, third-generation sequencing, and our devised bioinformatic pipeline (Fig. 3.1) allowed for an accurate, comprehensive sampling of mormyrid *epdl2* sequences, and would be easily extensible to other gene targets.

Multiplexed amplicon Nanopore sequencing is a viable and cost-effective option to study duplicated regions across divergent taxa, and we envision our approach as a valuable tool for researchers across many study systems.

While this Nanopore-based amplicon sequencing scheme was successful, we note that this approach was limited by how conserved sequences were outside both the start and stop codons on every paralog. Importantly, the number of *epdl2* paralogs we detected for each sample (Table 3.1) should not be considered definitive, particularly for difficult to amplify samples from post-duplications taxa (footnotes of Table 3.2 and Table 3.7). Definitive conclusions about copy number variation in *epdl2* will be greatly facilitated by future genome sequencing efforts in *Marcusenius ntemensis* and *Paramormyrops*, particularly the ~40 Kbp genomic region where these duplicates lie in tandem.

Paramormyrops underwent three *epdl2* duplications

We obtained a phylogenetically comprehensive *epdl2* gene tree that unambiguously indicates that *epdl2* duplications are not widely spread in Mormyridae, but rather confined to a recent but speciose branch centered around *Paramormyrops* and likely restricted to a large subset of these (Fig. 3.4). Whatever role *epdl2* plays in electric organs, its duplications were not necessary for the origin of electrogenesis in African weakly electric fish. From this gene tree, we infer the existence of four *epdl2* gene copies in *Paramormyrops*. Although sequence divergence between these paralogs is low, the four paralogs we discovered likely cover the standing paralog diversity. Each paralog group is supported by several, often distantly related *Paramormyrops* species, plus most of the detected sequences from duplicated lineages are clearly assigned to one paralog (Fig. 3.4). Our results support a duplication scenario that entails three duplication events, over an alternative with only two (Fig. 3.13), because the branching pattern of the *epdl2* gene tree (Fig. 3.4) fits better with the former scenario. In addition, we propose that all duplications occurred in a single lineage, the LABED; supported by the broad taxonomic range observed within each *epdl2* paralog cluster. A small variant to this idea is that the duplication that led to *epdl2.1* happened more recently, in an ancestor of *M. ntemensis* plus the “curvifrons” and

“cabrae” complexes (Fig. 3.5). However, *epd12.1* is the most divergent paralog in the *epd12* gene tree (Fig. 3.4). More likely, *epd12.1* could have been lost in an ancestor to the “magnostipes” complex plus its sister taxa (Fig. 3.5). Further potential paralog losses are *epd12.2* in the “curvifrons” complex and *epd12.4* in the “cabrae” complex. Although these three gene loss patterns fit very well with our data, they should be regarded as tentative until additional genome sequencing efforts are undertaken.

We detected three *epd12* paralogs in *M. ntemensis* but only one copy in *Paramormyrops szaboi*. One explanation for this finding is that *M. ntemensis* is a *Paramormyrops* species that diverged after *P. szaboi* and the LABED. This is counter to currently accepted phylogenetic relationships, which consider *M. ntemensis* as the sister taxon to *Paramormyrops* (Sullivan et al. 2002, 2004; Lavoué et al. 2003). The alternative explanation of parallel, independent duplications in *M. ntemensis* and *Paramormyrops* is not supported by the sequence similarity between the *epd12* paralogs in these taxa. Another hypothesis is that *epd12* paralogs may have been lost in *P. szaboi*. However, we note that the *P. szaboi epd12* gene was consistently placed as the sister sequence to all *epd12* paralogs. If this gene is an *epd12* paralog descendant from the LABED, we would expect it to cluster into its paralog group. This is in fact what we observe in both *P. offouensis* and *P. SN3*: we detected only one *epd12* gene in each species (Table 3.1), yet they clearly belong to *epd12.3* (Fig. 3.4). Moreover, the observed position of the *P. szaboi epd12* gene within the gene tree is congruent with this species basal taxonomic placement within *Paramormyrops* (Fig. 3.5). Ultimately, this observation will be resolved by future genome sequencing efforts in *M. ntemensis* and *P. szaboi*, or potentially by sequencing *epd12* genes in other basal *Paramormyrops*. Regardless of the outcome of these efforts, our broad evolutionary inferences remain valid regardless of the relationship of *M. ntemensis* and the other *Paramormyrops*.

Selection Tests

We investigated selection patterns at branch and site levels in the osteoglossiform *epd12* gene tree. The strength of selection, measured as higher ω values, has unambiguously increased in the duplicated lineages of *Paramormyrops* compared to the non-duplicated mormyrid lineages.

Furthermore, in two of the duplicated lineages, presumably between the second and third duplication, this increase has met the detection threshold for positive selection (Fig. 3.7). This, however, is likely a conservative estimate; aBSREL's statistical power in our dataset is limited by two technical reasons (Smith et al. 2015): i) this test requires multiple testing correction, hence we only tested a few branches; and ii) branch length directly affects statistical power, and some of the interrogated branches where we did not detect positive selection exhibit the lowest levels of divergence (paralog clusters *epd12.1* and *epd12.4*, Fig. 3.4).

Our site tests uncovered evidence of multiple sites in *epd12* under positive selection across Mormyridae (Fig. 3.8), and independently, also of sites where selection has intensified in the duplicated lineages (Fig. 3.9). We focused on the ten sites common to the two sets (Table 3.8), because we reason that these are the sites most likely targeted by selection after the *epd12* duplications. While we observe that some sites are supported by a substitution at a single sample (e.g. sites 105, 153) others experienced widespread, often paralog-specific substitutions (e.g. sites 106, 150) (Table 3.9).

Although the methodological nature of the MEME test (and similar site tests) hinders its ability to identify branch-site combinations subject to positive selection, it employs an exploratory procedure (the EBF) to suggest branches where a positively selected site may have experienced this selection (Murrell et al. 2012). It is important to note that these specific branch-site combinations should not be heavily relied on due to the exploratory nature of their statistical support, but also because paralog sampling is incomplete and topological inaccuracies are expected in the terminal branches of the *epd12* gene tree. That said, because we screened these ten sites for augmented selection in duplicated lineages, it is to be expected that the branches where they have experienced positive selection are post-duplication branches. Reassuringly, for these ten sites, all but one branch-site pair with EBF support (Table 3.8) are located in duplicated lineages (Fig. 3.10, green bar). Since the two other flagged branches

for site 176 belong to the duplicated lineages (Fig. 3.10), this lone case is likely an example where positive selection could have acted on sites in the *epdl2* single-copy mormyrid lineages.

Taken together, our branch and site tests converge on the conclusion that the *epdl2* paralogs have experienced significant positive selection.

Functional Consequences of Molecular Evolution in *epdl2* Paralogs

EPDR proteins are widespread across eukaryotic groups, where they participate in a diverse gamut of often lineage specific functions. Given their low amino acid conservation, these functions are likely accomplished through shared structural and biochemical properties: a signal peptide, N-glycosylation sites, disulfide bonds, dimerization, and a hydrophobic pocket. This framework suggests that comparisons between the 3D structures of Epdl2 and Epdr1 facilitate inferences about expected broad functional features of Epdl2, despite low sequence conservation between them; different cellular localizations, since Epdr1 is regarded as a lysosomal protein (Park et al. 2019; Wei et al. 2019) whereas fish EPDRs are secreted into the extracellular space; and that detailed knowledge of Epdr1 function is still undetermined.

The structural annotation of *Paramormyrops szaboi* Epdl2 (Fig. 3.11) and its 3D model prediction (Fig. 3.12) corroborate that our methodological approach successfully amplified, sequenced, and identified functional *epdl2* genes. As predicted, this gene has a signal peptide on its N terminus and two N-glycosylation sites whose position is conserved in fish EPDRs (Müller-Schmid et al. 1992; Ortí and Meyer 1996; Suárez-Castillo and García-Arrarás 2007). Although these positions are not conserved in other EPDRs including Epdr1, glycosylation in the latter may be necessary for its function (Wei et al. 2019). Our 3D model suggests that these sites are exposed in Epdl2, and therefore likely accessible to the glycosylation machinery. Also predicted is a very high spatial similarity between Epdl2 and Epdr1, the only ependymin protein with an elucidated crystal structure. The Epdl2 3D model showcases the structural hallmarks of EPDRs: conserved disulfide bonds, β strands that form two antiparallel β sheets connected by a linker region, and α helixes near the C terminus (McDougall et al. 2018; Park et al. 2019;

Wei et al. 2019). The β sheets have been predicted to form a deep hydrophobic pocket or groove in all EPDRs with potential roles in ligand-binding (McDougall et al. 2018), and its existence and functionality have been demonstrated in Epdr1, including the linker region as an operational element of this pocket (the lid) (Park et al. 2019; Wei et al. 2019). Our model includes all these components arranged strikingly similar to Epdr1 and a highly conserved proline residue located to this region, as predicted by (McDougall et al. 2018). Every structural homology unambiguously identifies the location of this hydrophobic pocket in EpdI2 (Fig. 3.12D, see description in Results).

The Epdr1 crystal structure contains a β -sheet hydrophobic pocket similar to bacterial proteins of the LolA superfamily (Park et al. 2019; Wei et al. 2019), which participate in widespread functions through binding diverse hydrophobic ligands in their pocket (Wei et al. 2019). By way of its hydrophobic pocket, Epdr1 has been proposed to bind hydrophobic molecules in general and potentially sequester digested lipids in lysosomes (Park et al. 2019), and to participate in lipid transport and/or metabolism and likely be a lysosomal activator (Wei et al. 2019). The specificity of EPDR ligands likely depends on the amino acid residues in these proteins' hydrophobic pocket: sequence conservation in this region is low not only among EPDRs, but also between Epdr1 and LolA (Wei et al. 2019). This apparent versatility in their binding affinities could relate to EPDRs' involvement in a remarkable wide range of functions (Shashoua 1985; Schmidt et al. 1995; Pradel et al. 1999; Jackson et al. 2006; Zheng et al. 2006; Staats et al. 2016; Hall et al. 2017). It is straightforward to conceive that duplicated EPDR copies could be tailored to slightly different ligands through changes in their amino acid sequences at key positions. All this evidence underscores the functional relevance of the shape of the hydrophobic pocket in EPDR proteins. We put forward functional predictions for the ten sites of highest interest (Table 3.8) based on their spatial arrangement in our EpdI2 structural model and on the elucidated 3D conformation of Epdr1. Sites 105 and 106 are located in the hydrophobic pocket's lid, and sites 150, 153, 172, and 176 map to

two β strands on the same β sheet in the pocket's lining (Fig. 3.12E-F). However, of the latter, only site 153 is oriented towards the inside of the pocket, spatially close to the conserved proline site.

Additionally, the crystal structures of Epdr1 demonstrates that two Epdr1 chains associate into a homodimer through contacts between one of the β sheets and a linker region (Park et al. 2019; Wei et al. 2019). Given the predicted structural similarity between Epdl2 and Epdr1, we expect that Epdl2 also forms homodimers mediated by its structural homologs. Of our ten sites of interest, sites 150, 153, 172, and 176 belong in the β sheet that mediates dimer formation, and three of these sites, 150, 172, and 176, have side chains pointing towards the dimerization surface. In addition, sites 125, 126, 127, and 129 all locate to the linker region known to be an important component of the contacts between the monomeric subunits (Fig. 3.12E-F).

We note that four of our sites of interest map to the β sheet that lines part of the hydrophobic pocket and also participates in the dimer conformation. This complicates deductions about each of these site's biological function. Nevertheless, we find it encouraging that our results from the selection tests and the inferences from the structural predictions blend well with each other. Collectively, they suggest selection-driven diversification in amino acid residues that coherently affects protein function. We emphasize that the procedure we used to identify signatures of positive selection relies on coding sequences and is therefore uninformed by proteic structures, and the 3D model prediction is not aware of evolutionary interesting sites. We speculate that three of the ten sites of highest interest (105, 106, and 153) may participate in the protein's affinity for a ligand, whereas the remaining seven sites (125, 126, 127, 129, 150, 172, and 176) may influence homodimer formation. We note that a necessary step towards efficient subfunctionalization in duplicated Epdl2 proteins is the correct pairing of the Epdl2 paralog chains into homodimers, which is likely to be achieved by paralog-specific amino acid changes in the dimerization surface.

Conclusions

In this analysis, we sequenced and annotated the three *epdl2* paralogs present in *Paramormyrops kingsleyae* and developed a novel, Nanopore-based, sequencing and bioinformatics approach to analyze amplicons from duplicated targets, which yielded *epdl2* sequences from 20 mormyrid taxa. We propose that as many as four *epdl2* paralogs resulted from three tandem duplications events in an early *Paramormyrops* ancestor, and we demonstrate that these *epdl2* paralogs, and specific sites in them, have experienced increased selection rates and have been targets of positive selection. Finally, we identify that these specific sites are located in protein regions relevant to ligand binding and homodimer formation. Presently, the specific functional role of Epd12 in mormyrid electrocytes is unknown. The homodimeric configuration of Epd1, and hence the expected arrangement in Epd12, contains both hydrophobic pockets next to each other, open towards a relatively flat surface on the dimer. It has been proposed that this allows membrane binding “with an extensive contact surface for the binding and possible extraction and solubilization of target lipids” (Wei et al. 2019). We hypothesize based on these proposed functions that Epd12 may be related to the shaping/maintenance of the plasma membrane via selectively removing specific phospholipids. Selective removal of membrane lipids could change their composition and distribution, and therefore influence the membrane’s biochemical properties like shape, curvature, electrical charges, fluidity, and local protein composition. Our hypothesis makes general predictions about Epd12 expression and localization in mormyrids. Epd12, or some of its paralogs, is expected to be more abundant in electrocytes with more complex plasma membrane configurations, like papillae or penetrations. Likewise, Epd12 would be more commonly found bound to the external face of the plasma membrane, and this surface’s lipid composition would have relatively low levels of Epd12’s targets. Given the well-established connection between the electrocytes’ membrane traits and the resulting EOD (Bennett and Grundfest 1961; Szabo 1961; Bennett 1971; Hopkins 1999; Gallant et al. 2011; Carlson and Gallant 2013), increased plasticity

and control of this membrane's biophysical properties could facilitate EOD divergence, which may contribute to species diversification.

APPENDIX

Table 3.1. Source of *epdl2* sequences, number and size of *epdl2* genes found in every species studied

species	source of <i>epdl2</i> sequence(s)	<i>epdl2</i> genes found	gene sizes (bp) ¹
<i>Brevimyrus niger</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1823
<i>Brienomyrus brachyistius</i>	genome assembly available	1	2166
<i>Campylomormyrus sp</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1801
<i>Gnathonemus petersii</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1779
<i>Gymnarchus niloticus</i>	genome assembly available (NCBI bioproject PRJNA423259)	1	5280
<i>Ivindomyrus marchei</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	discarded ²	-
<i>Marcusenius moori</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1808
<i>Marcusenius ntemensis</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	3	1675, 1806, 1823
<i>Mormyrops zanclostris</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1869
<i>Mormyrus rume</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1816
<i>Paramormyrops curvifrons</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	3	1676, 1819, 1823
<i>Paramormyrops hopkinsi</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	3	1793, 1821, 1824
<i>Paramormyrops kingsleyae (biphasic)</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	3	1675, 1810, 1813
<i>Paramormyrops kingsleyae (triphasic)</i>	paralog-specific PCR amplification & Sanger-sequencing	3	1675, 1809, 1814
<i>Paramormyrops magnostipes Type I</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	3	1815, 1823, 1825
<i>Paramormyrops magnostipes Type II</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	2	1817, 1825

Table 3.1 (cont'd)

<i>Paramormyrops ngounensis</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	2	1676, 1815
<i>Paramormyrops offouensis</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1823
<i>Paramormyrops SN2</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	discarded ³	
<i>Paramormyrops SN3</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1824
<i>Paramormyrops SN9</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	discarded ³	
<i>Paramormyrops szaboi</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1820
<i>Paramormyrops tenuicauda</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	3	1805, 1817, 1819
<i>Paramormyrops teugelsi</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	discarded ³	
<i>Petrocephalus simus</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1792
<i>Pollimyrus sp</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1551
<i>Scleropages formosus</i>	genome assembly available (NCBI GeneID 108934263)	1	2918
<i>Stomatorhinus ivindoensis</i>	DNA extraction, PCR amplification & multiplexed amplicon sequencing with ONT	1	1808

¹From start to stop codon. bp = base pairs

²This sample was likely an incorrectly identified *P. kingsleyae*

³No *epd2* genes were successfully amplified and sequenced

Table 3.2. Primers used to amplify and sequence *epdl2* paralogs across *Mormyridae*

primer	sequence (5'-3')	target gene	target clade	PCR annealing temperature (°C)
<i>epdl2.1_1F</i>	CAGCCAGTGCCTCTACCATT TGC	<i>epdl2.1</i>	<i>P. kingsleyae</i>	71
<i>epdl2.1_1R</i>	AGGAATGAAACGAACAAAA GTTCAGGCAAGT			
<i>epdl2.2_2F</i>	GGTGAAGTGCAGGTCTAGT TTG	<i>epdl2.2</i>	<i>P. kingsleyae</i>	65
<i>epdl2.2_2R</i>	ACAGAAGTTCAGGCAACTT TAACCTC			
<i>epdl2.3_2F</i>	AACCTACAAGGGACTTTGCT AACCC	<i>epdl2.3</i>	<i>P. kingsleyae</i>	64
<i>epdl2.3_1R</i>	GCCATGGACTACTTCTACAG CGCAG			
<i>epdl2_Sange r_1</i>	TTCCTATCTGCCCTGGTA	<i>epdl2.1, epdl2.2, epdl2.3</i>	<i>P. kingsleyae</i>	NA
<i>epdl2_Sange r_2</i>	TATCGCTGGGATTTTCGAG	<i>epdl2.1, epdl2.2, epdl2.3</i>	<i>P. kingsleyae</i>	NA
<i>epdl2_Sange r_3</i>	CTGTTATCCTAGGGATGAG G	<i>epdl2.1, epdl2.2, epdl2.3</i>	<i>P. kingsleyae</i>	NA
<i>epdl2_Sange r_4</i>	CTGTCCAGGTTCTAATGC	<i>epdl2.1, epdl2.2, epdl2.3</i>	<i>P. kingsleyae</i>	NA
<i>epdl2_Sange r_5</i>	TTCCAGACAAGCTCACTG	<i>epdl2.1, epdl2.2</i>	<i>P. kingsleyae</i>	NA
<i>epdl2_0408_ F01</i>	AGCARCRACACATTTTTGK	all <i>epdl2</i> paralogs	<i>Ivindomyrus, Marcusenius ntemensis, Mormyrus, Paramormyrops</i>	60 ¹
<i>epdl2_0408_ R01</i>	AGGGTTTSGAGTCAGGR			
<i>epdl2_Pol+St o_F01</i>	GTTGTTTTCAAAGTCGTCC	all <i>epdl2</i> paralogs	<i>Pollimyrus, Stomatorhinus</i>	60
<i>epdl2_Pol+St o_R01</i>	TGAACCTTGGATTCACAC			
<i>epdl2_Bre+H yp_F01</i>	TTACTTTCAGTGCTGTATC	all <i>epdl2</i> paralogs	<i>Brevimyrus</i>	53
<i>epdl2_Bre+H yp_R01</i>	CAGAGAATGCAGATAATTC AC			
<i>epdl2_Mar+C am_F01</i>	GTTGTTTTCAAAGTCGTCC	all <i>epdl2</i> paralogs	<i>Campylomormy rus, Gnathonemus, Marcusenius</i>	52, 53, 55 ²
<i>epdl2_Mar+C am_R01</i>	TCTCTCTCCCTCTGATATA			

Table 3.2 (cont'd)

<i>epdl2</i> _Morps_F01	CGAATCCTTAAATCCCAATC	all <i>epdl2</i> paralogs	<i>Mormyrops</i>	55
<i>epdl2</i> _Morps_R01	CTGTAACGATTACATGAC			
<i>epdl2</i> _Pet_F01	AGGGACAAYTTAGTCAGGA	all <i>epdl2</i> paralogs	<i>Petrocephalus</i>	60
<i>epdl2</i> _Pet_R01	TTGAGTCAGAGRACACAGT			

NA: primer not used in PCR reactions

¹Touch-up annealing temperatures for: *Paramormyrops curvifrons*, *Paramormyrops offouensis*, *Paramormyrops* SN2, *Paramormyrops* SN9, *Paramormyrops teugelsi*. Touch-up conditions: one set was comprised of 7 cycles from 57 to 60°C in 0.5°C increments. Each PCR consisted of 4 sets followed by 4 additional cycles at 60°C.

²*Campylomormyrus*: 52°C, *Gnathonemus*: 55°C, *Marcusenius*: 53 °C

Table 3.3. PCR reagents and final concentrations used to amplify each *epdl2* paralog in *P. kingsleyae*

Reagent	Final concentration
5x Q5 reaction buffer	1x
dNTPs	200 µM each dNTP
Forward and Reverse Primers	1.0 µM each
Template DNA	<1000 ng
Q5 polymerase	0.02 U/µl
5x GC enhancer for Q5	0.5x

Table 3.4. PCR conditions used to amplify each *epdl2* paralog in *P. kingsleyae*

Step	temperature (°C)	time
Preheat lid + block	98	-
Initial denaturation	98	30 s
30 cycles of:		
denaturation	98	10 s
annealing	primer-dependent (Table 3.2)	25 s
extension	72	2 min
Final extension	72	2 min

Table 3.5. Mormyrid species and their NBCI bioproject codes that guided *epdl2* primer design

species	BioProject
<i>Boulengeromyrus knoepffleri</i>	PRJNA526756
<i>Brevimyrus niger</i>	PRJNA526749
<i>Genyomyrus donnyi</i>	PRJNA529465
<i>Gnathonemus echidnorhynchus</i>	PRJNA529468
<i>Hyperopisus bebe</i>	PRJNA529477
<i>Isichthys henryi</i>	PRJNA529470
<i>Ivindomyrus marchei</i>	PRJNA529476
<i>Marcusenius schilthuisiae</i>	PRJNA529469
<i>Mormyrops attenuatus</i>	PRJNA530793
<i>Mormyrops boulengeri</i>	PRJNA530782
<i>Mormyrops zancloirostris</i>	PRJNA530797
<i>Mormyrus hasselquistii</i>	PRJNA542939
<i>Mormyrus iriodes</i>	PRJNA542943
<i>Mormyrus probosciostris</i>	PRJNA530791
<i>Myomyrus macrops</i>	PRJNA423275
<i>Myomyrus pharao</i>	PRJNA547756
<i>Paramormyrops hopkinsi</i>	PRJNA547741
<i>Paramormyrops magnostipes</i>	PRJNA547743
<i>Petrocephalus microphthalmus</i>	PRJNA423286
<i>Petrocephalus schoutedeni</i>	PRJNA547742
<i>Petrocephalus sullivanii</i>	PRJNA427158
<i>Petrocephalus zakoni</i>	PRJNA547751
<i>Pollimyrus isidori</i>	PRJNA547785
<i>Pollimyrus plagiostoma</i>	PRJNA547754
<i>Stomatorhinus walkeri</i>	PRJNA547748

Table 3.6. PCR reagents and final concentrations used to amplify all *epd12* paralogs across *Mormyridae*

Reagent	Final concentration
5x Q5 reaction buffer	1x
dNTPs	200 μ M each dNTP
Forward and Reverse Primers	0.9 μ M each ¹ or 0.5 μ M each ²
Template DNA	<1000 ng
Q5 polymerase	0.02 U/ μ l
5x GC enhancer for Q5	0.1x

¹*epd12_0408* primers with *Ivindomyrus marchei*, *Marcusenius ntemensis*, and all *Paramormyrops* spp

²*epd12_0408* primers with *Mormyrus rume*, and all other primers

Table 3.7. PCR conditions used to amplify all *epd12* paralogs across *Mormyridae*

Step	temperature (°C)	time
Preheat lid + block	98	-
Initial denaturation	98	30 s
30 ¹ cycles of:		
denaturation	98	10 s
annealing	primer- and species-dependent (Table 3.2)	20 s ²
extension	72	75 s ³
Final extension	72	2 min

¹32 cycles for: *Brevimyrus niger*, *Campylomormyrus* sp, *Ivindomyrus marchei*, *Marcusenius ntemensis*, *Paramormyrops curvifrons*, *Paramormyrops offouensis*, *Paramormyrops* SN2, *Paramormyrops* SN9, *Paramormyrops teugelsi*

²25 seconds for: *Brevimyrus niger*, *Campylomormyrus* sp, *Ivindomyrus marchei*, *Marcusenius ntemensis*

³90 seconds for: *Brevimyrus niger*, *Campylomormyrus* sp

Table 3.8. Sites along *epdl2* that have experienced positive selection in the osteoglossiform *epdl2* gene tree and have evolved at increased ω rates in duplicated vs non-duplicated mormyrid lineages. Branches supported by each EBF value are marked by rectangles in Figure 3.10 (EBF values for a given site are arranged to match their corresponding branches from top to bottom).

Site	Estimated number of branches under positive selection	Empirical Bayes Factor (EBF)
105	1	3.7×10^{11}
106	0	
125	1	5.1×10^4
126	2	1623, 1×10^{26}
127	2	2294, 3.2×10^{10}
129	2	589, 1×10^{26}
150	2	2000, 234
153	1	8.6×10^{12}
172	4	2.4×10^4 , 2.4×10^4 , 1×10^{26} , 1028
176	3	1298, 3409, 1.0×10^4

Table 3.9. Summary of the amino acid substitutions observed in the *Epd12* paralogs at the ten sites under positive selection and increased ω rates in duplicated vs non-duplicated mormyrid lineages

site	ancestral amino acid	derived amino acid residues observed in each paralog (total)			
		<i>epdl2.1</i> (5)	<i>epdl2.4</i> (6)	<i>epdl2.2</i> (6)	<i>epdl2.3</i> (10)
105	F	-	-	S (1)	-
106	P	R (5)	L (6)	R (6)	R (10)
125	S	-	N (1)	N (6)	N (10)
126	S	-	-	-	L (8)
127	A	-	-	D (6)	-
129	G	-	S (1)	S (2)	S (2)
150	Q	L (5)	L (4), K (2)	K (6)	K (10)
153	F	-	-	C (1)	-
172	L	-	-	R (1)	W (7), R (2)
176	S	-	C (1)	-	N (7)

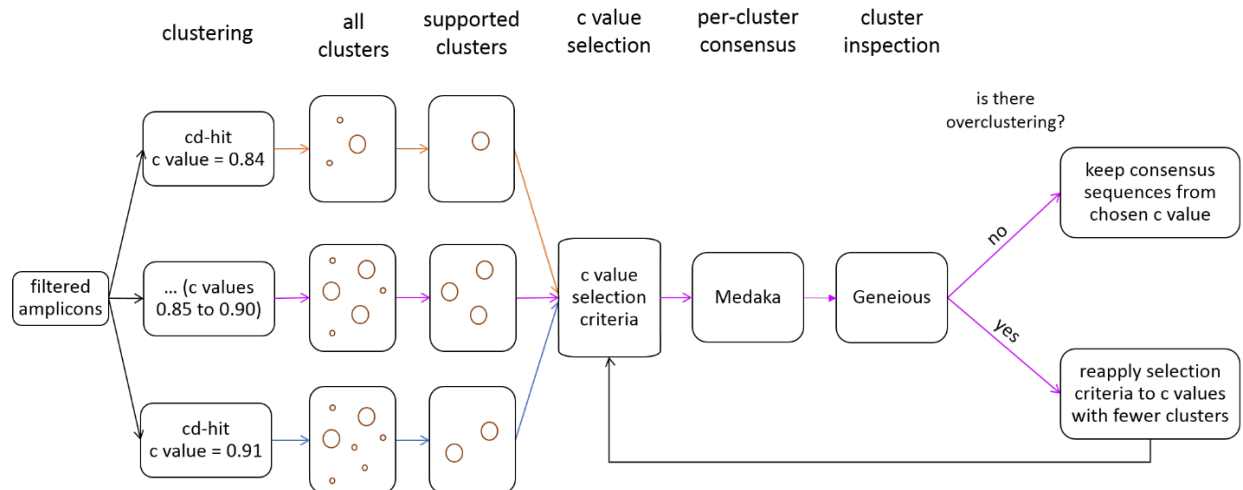


Figure 3.1. Graphical summary of the bioinformatic pipeline we leveraged to identify *epd12* genes in the filtered amplicons from each species, see Methods for details. Same-colored connector arrows represent the analysis with a specific value for *c* (cd-hit's sequence identity threshold parameter, range analyzed 0.84-0.91). Magenta connector arrows represent the analysis with the *c* value chosen with the selection criteria.

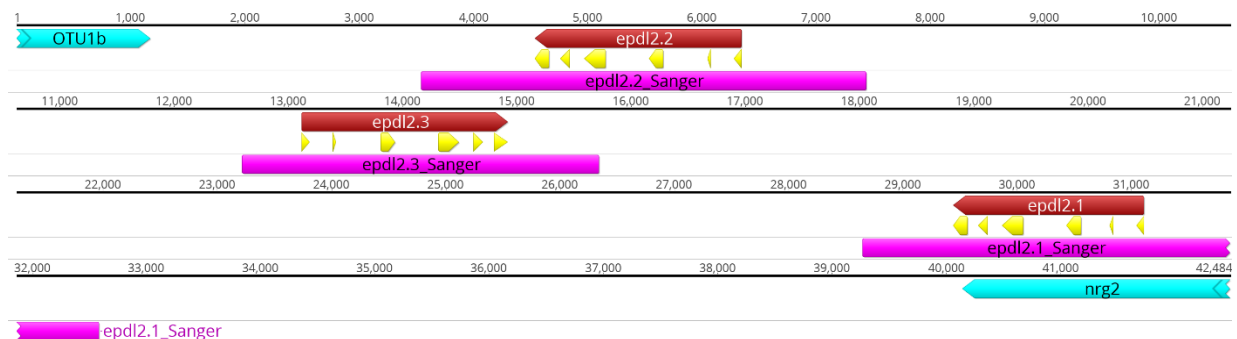


Figure 3.2. Genomic region with the *epd12* paralogs in *Paramormyrops kingsleyae*. DNA sequence is represented by the black line, numbers above indicate base positions, annotations are shown below the line. Flanking genes are depicted in cyan, *epd12* paralogs in red (start to stop codon) and yellow (CDSs). Arrows direction indicate gene membership to a strand (right for +, left for -). Magenta blocks mark the Sanger-sequenced regions.

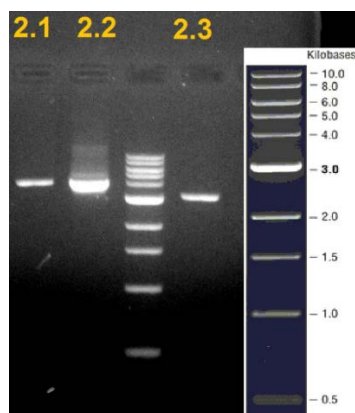


Figure 3.3. Cleaned PCR products for each *Paramormyrops kingsleyae epd12* paralog. Lanes with a PCR product are labeled with the respective abbreviated paralog name. The unlabeled lane (lane 3) holds a DNA ladder, its size legend is shown inserted

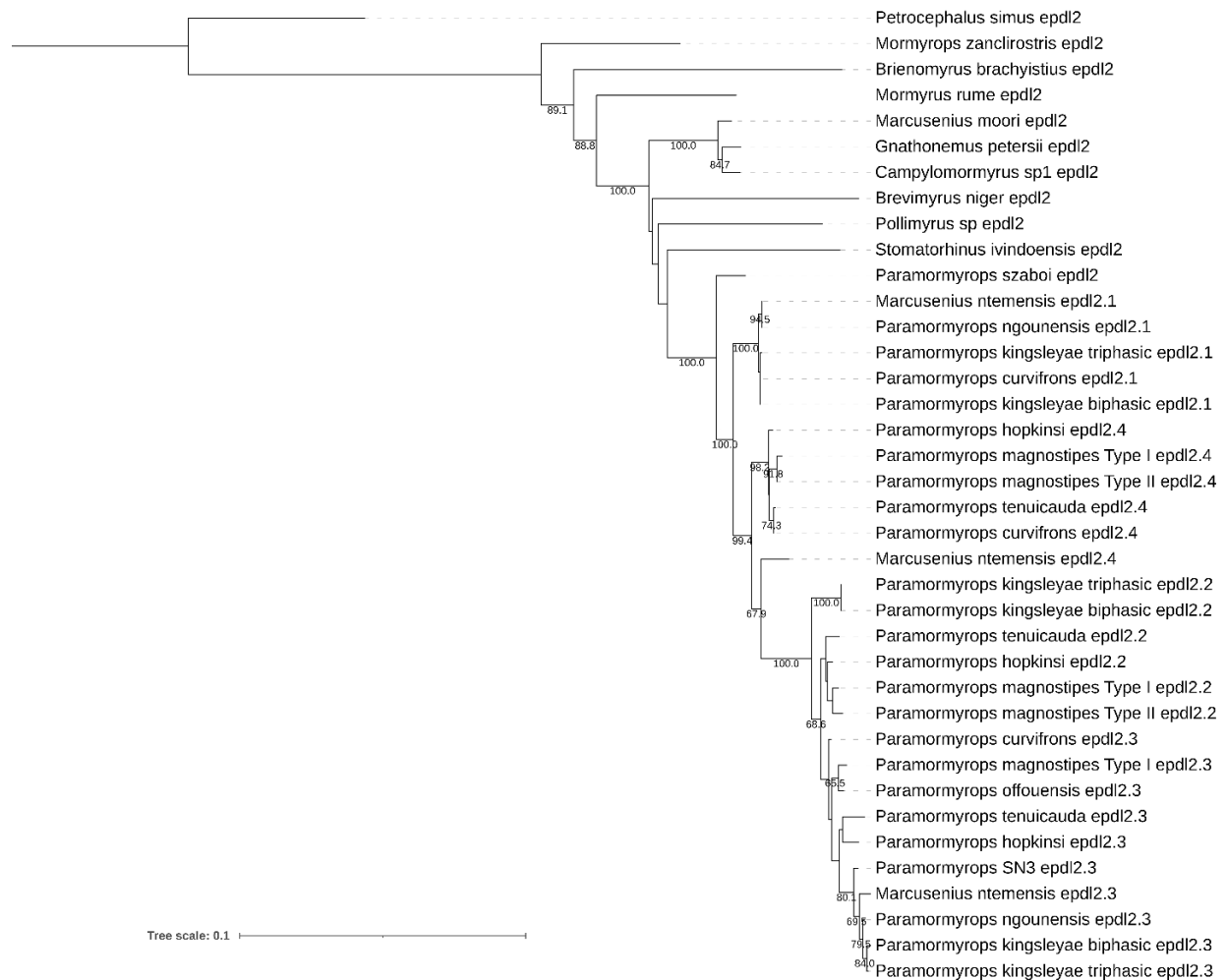


Figure 3.4. Mormyrid *epd12* gene tree. Bootstrap support values > 60% are shown. *epd12* paralogs are classified based on our best hypothesis of *epd12* duplications

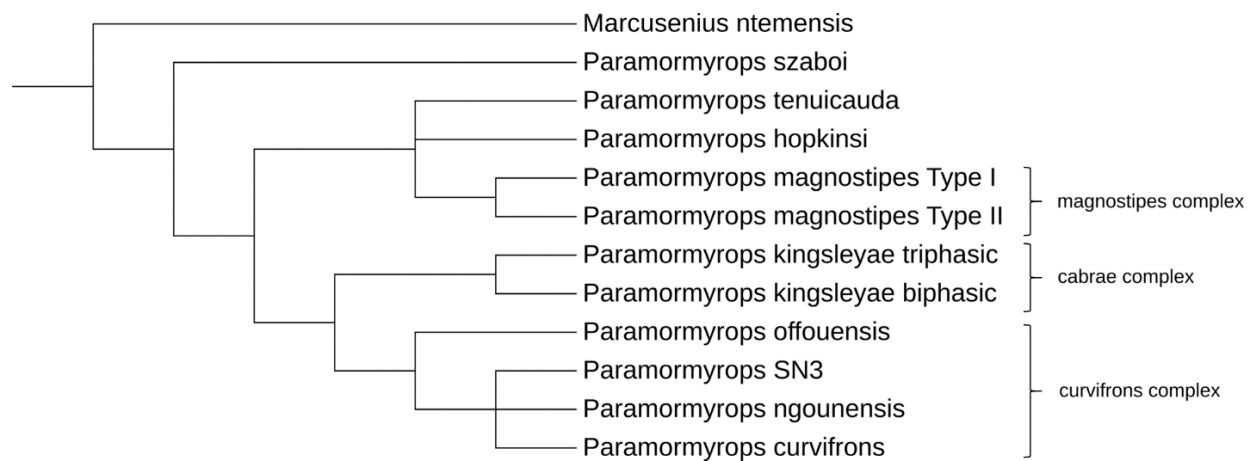


Figure 3.5. Broad phylogenetic relationships predicted among the studied species from the *Paramormyrops* + *Marcusenius ntemensis* clade. Based on the AFLP phylogeny from (Sullivan et al. 2004)

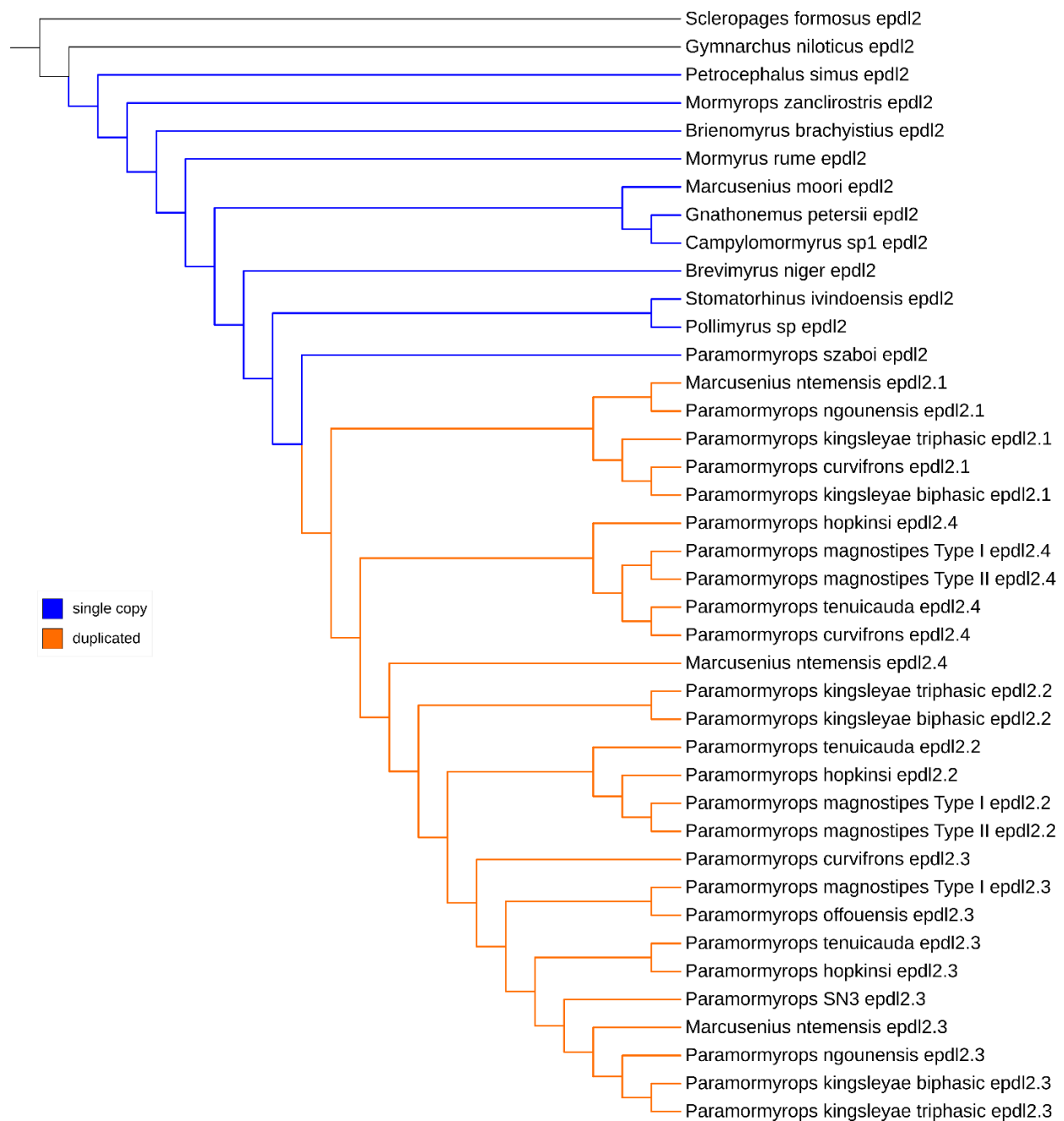


Figure 3.6. Topology of the *epl2* gene tree with duplicated and non-duplicated (single copy) mormyrid lineages indicated. This branch partition was used in the RELAX and Contrast-FEL tests.

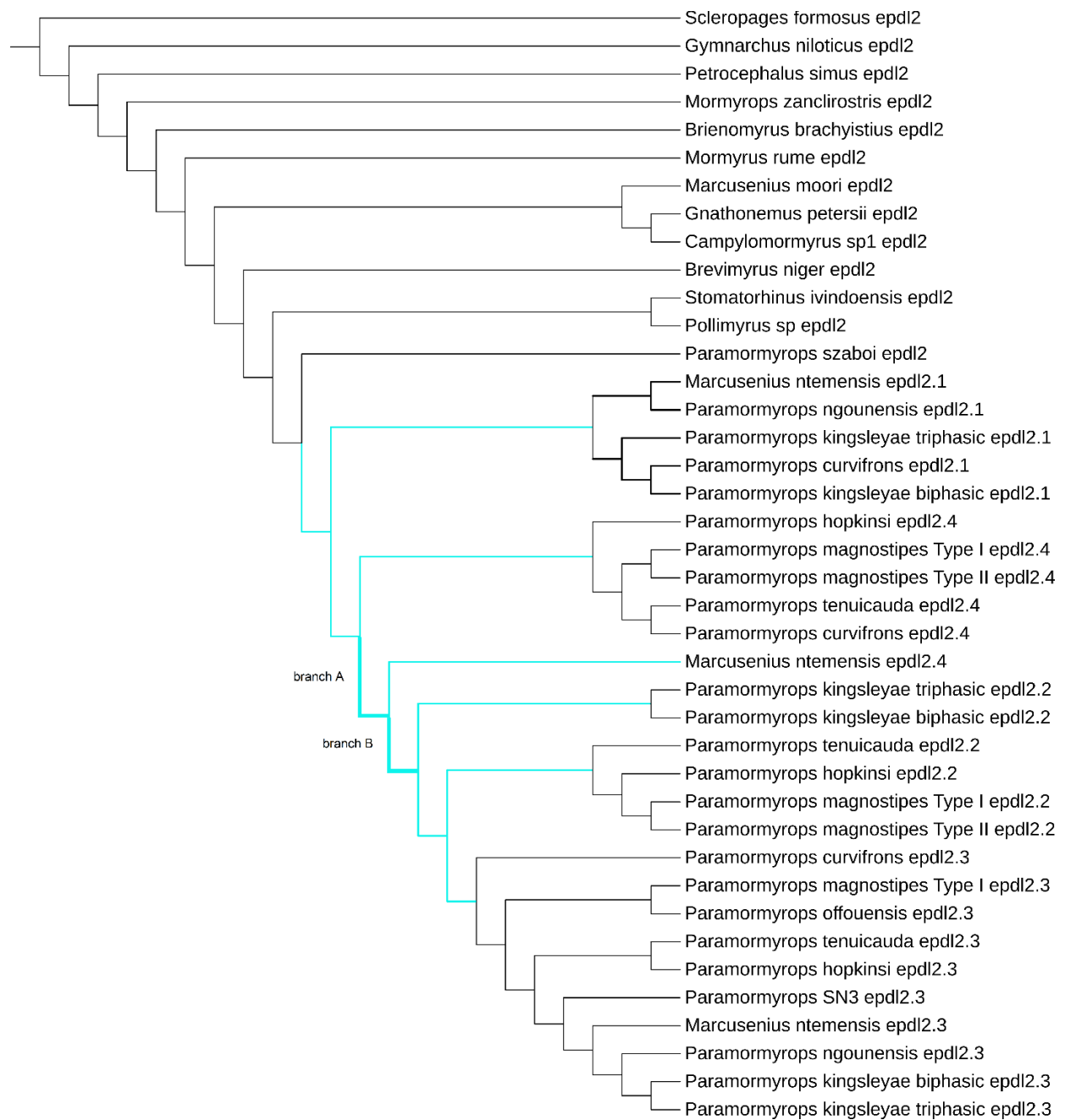


Figure 3.7. Topology of the *epl2* gene tree with branches tested for positive selection with aBSREL indicated. Branches with significant results are labeled A and B.

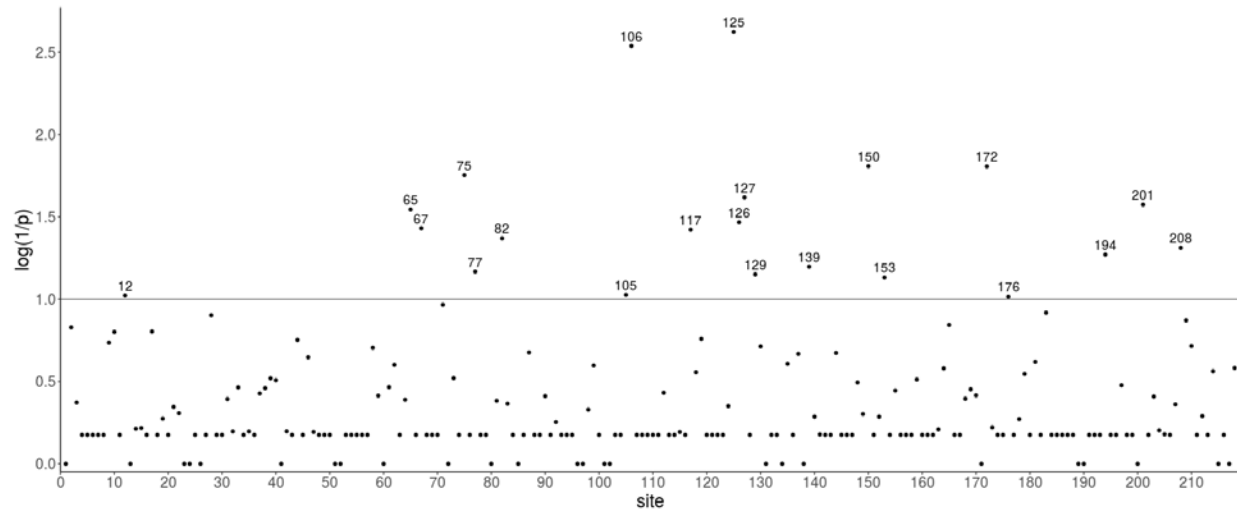


Figure 3.8. Sites along *epdl2* that have experienced positive selection (MEME, $p < 0.1$) in the osteoglossiform taxa studied. p values have been transformed so that higher values on the y axis represent lower p values. Horizontal line marks the significance threshold and significant sites have been labeled.

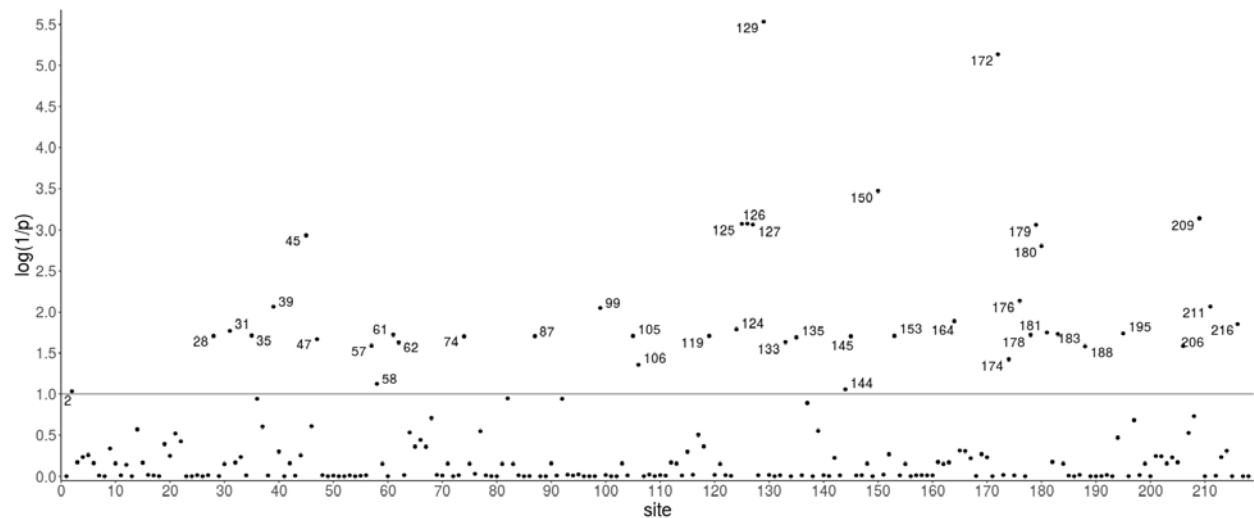


Figure 3.9. Sites along *epdl2* with higher ω values in duplicated vs non-duplicated mormyrid lineages (Contrast-FEL, $p < 0.1$). p values have been transformed so that higher values on the y axis represent lower p values. Horizontal line marks the significance threshold and significant sites have been labeled.

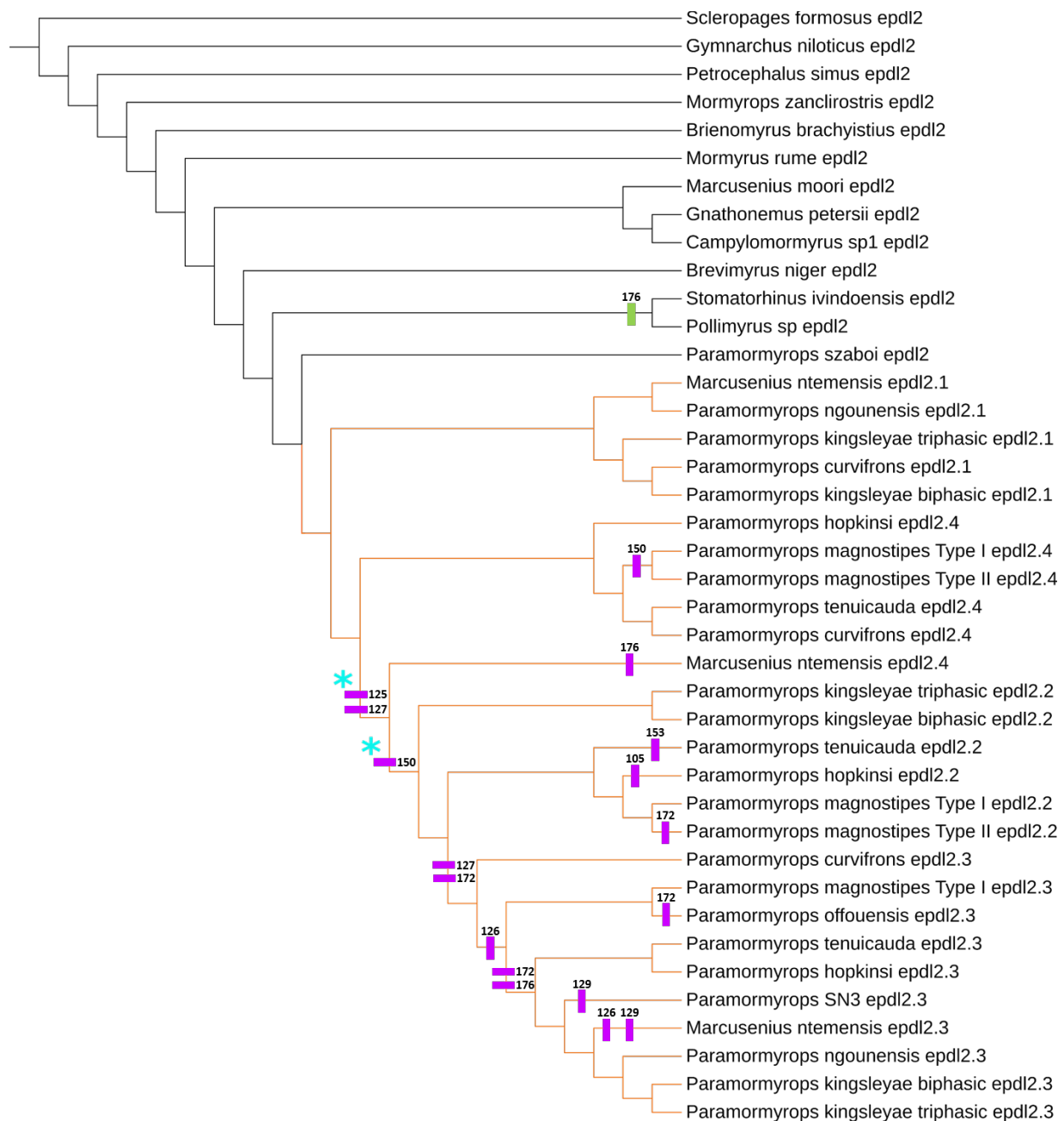


Figure 3.10. Topology of the *epd12* gene tree, we highlight branches and sites where we detected signals of selection. Orange branches are recently duplicated *epd12* paralogs, selection has intensified in these branches relative to the non-duplicated mormyrid lineages. Branches with asterisks experienced positive selection. Colored rectangles are labeled with sites along *epd12* where selection has intensified in duplicated lineages and have experienced positive selection. These rectangles are placed on the branches where exploratory evidence suggests they underwent positive selection. All rectangles (purple) but one (green) map to the duplicated lineages.

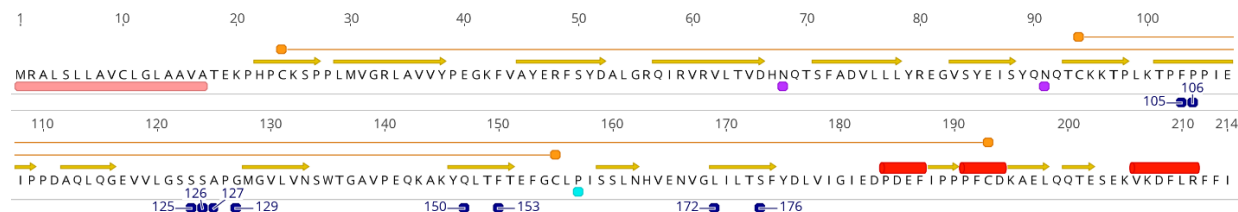


Figure 3.11. *Paramormyrops szaboi* *Epd12* predicted amino acid sequence, annotated with salient structural properties. Pink: signal peptide; orange: conserved cysteine residues forming disulfide bonds; purple: N-glycosylation sites; blue: positively selected sites with increased ω rates in *epd12* paralogs, number labels indicate sites' position in the osteoglossiform alignment of *epd12* CDSs. Yellow arrows and red blocks are β strands and α helices interpreted by Geneious from the 3D model

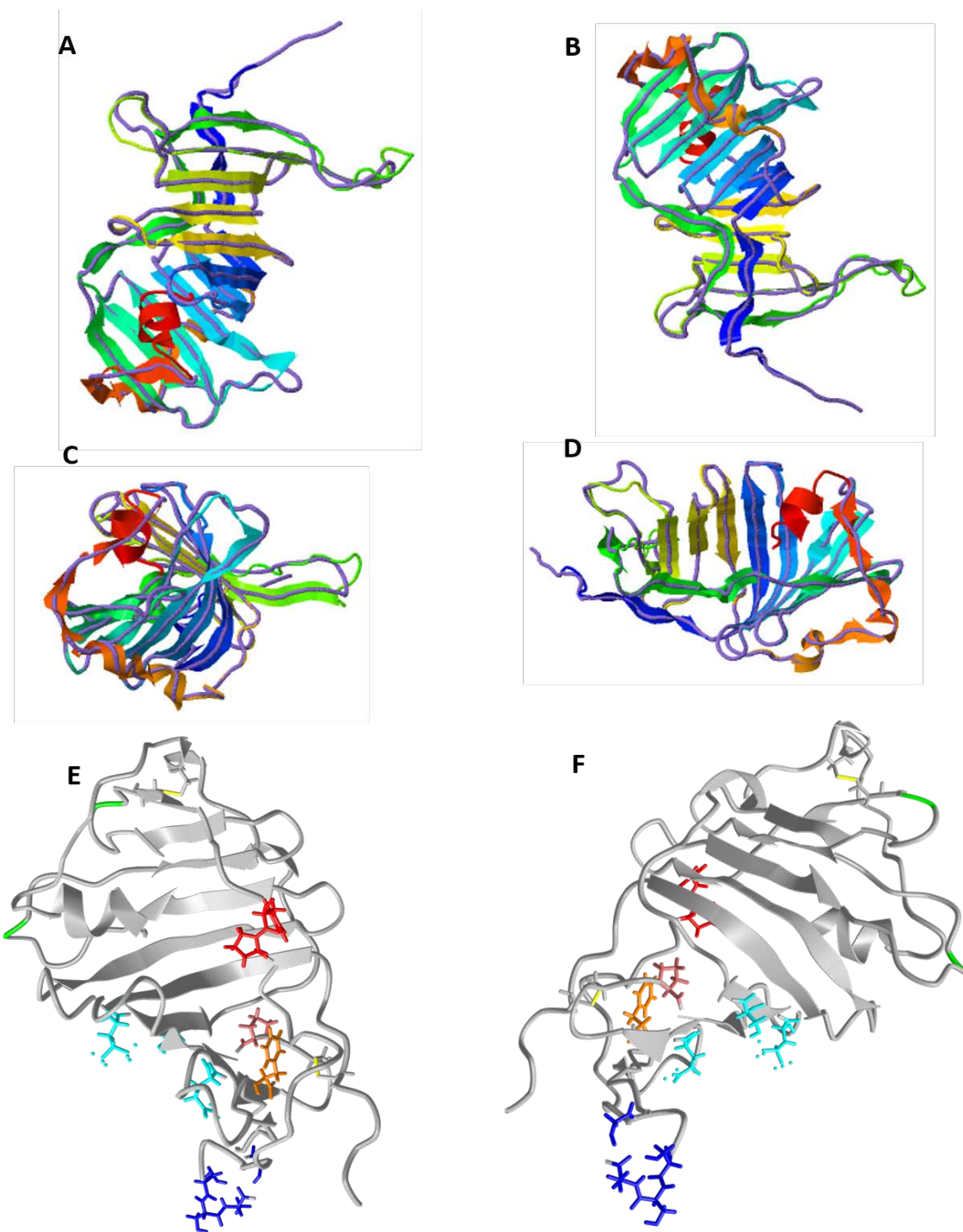


Figure 3.12. Predicted 3D structure of *Paramormyrops szaboi* Epd12. A-D: four different views superimposed on its best structural analog, *Xenopus tropicalis* epdr1 (purple backbone trace). E-F: two different views highlighting the location of structural features (disulfide bonds yellow, N-glycosylation sites green) and residues of interest (colored sticks): conserved proline residue pink, and our ten sites of highest interest (105-106 red, 125-7, 129 blue, 150, 172, 176 cyan, 153 orange). Pink, red and orange residues are likely oriented towards the hydrophobic pocket (E), blue and cyan residues are probably directed towards the exterior (F)

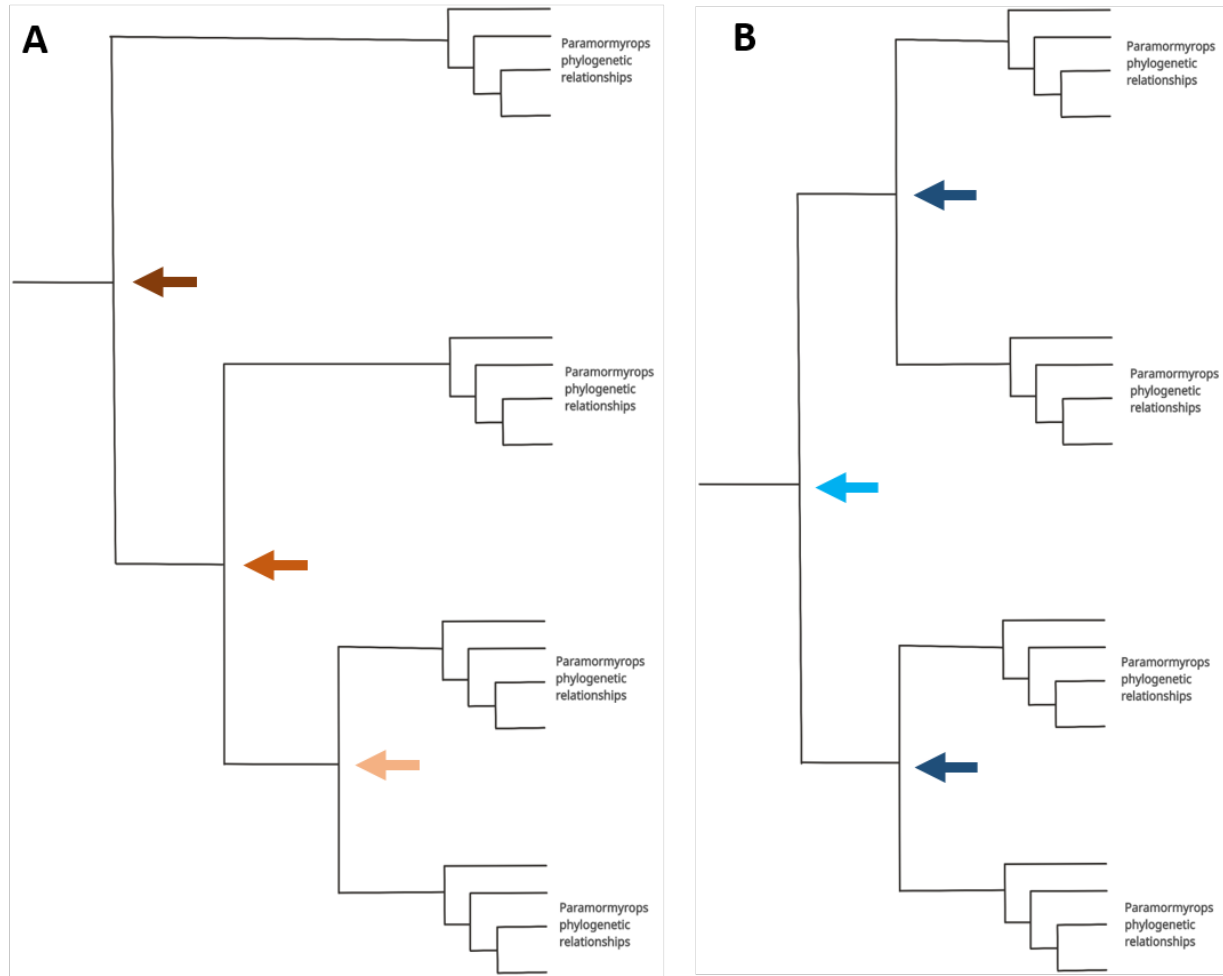


Figure 3.13. Theoretical gene trees with the topological patterns expected under two duplication scenarios. Arrows indicate gene duplications, same-colored arrows mark gene duplications stemming from the same duplication event. A: Three duplications scenario, each duplication yields one additional paralog. B: two duplications scenario, each duplication doubles the number of paralogs.

REFERENCES

REFERENCES

- Abascal, F., R. Zardoya, and M. J. Telford. 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38:7–13.
- Albertson, R. C., K. E. Powder, Y. Hu, K. P. Coyle, R. B. Roberts, and K. J. Parsons. 2014. Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. *Mol. Ecol.* 23:5135–5150.
- Almagro Armenteros, J. J., K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne, and H. Nielsen. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37:420–423.
- Alves-Gomes, J., and C. D. Hopkins. 1997. Molecular insights into the phylogeny of mormyrid fishes and the evolution of their electric organs. *Brain Behav. Evol.* 49:324–350.
- Arnegard, M. E., S. M. Bogdanowicz, and C. D. Hopkins. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. *Evolution* 59:324–343.
- Arnegard, M. E., B. S. Jackson, and C. D. Hopkins. 2006. Time-domain signal divergence and discrimination without receptor modification in sympatric morphs of electric fishes. *J. Exp. Biol.* 209:2182–2198.
- Arnegard, M. E., P. B. McIntyre, L. J. Harmon, M. L. Zelditch, W. G. R. Crampton, J. K. Davis, J. P. Sullivan, S. Lavoué, and C. D. Hopkins. 2010. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *Am. Nat.* 176:335–356.
- Bass, A. H. 1986. Species differences in electric organs of mormyrids: Substrates for species-typical electric organ discharge waveforms. *J. Comp. Neurol.* 244:313–330.
- Bennett, M. V. L. 1971. Electric Organs. Pp. 347–491 in W. S. Hoar and D. J. Randall, eds. *Fish Physiology*. Academic Press, London.
- Bennett, M. V. L., and H. Grundfest. 1961. Studies on the morphology and electrophysiology of electric organs. III. Electrophysiology of electric organs in mormyrids. Pp. 113–35 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, Amsterdam.
- Carlson, B. A., and J. R. Gallant. 2013. From sequence to spike to spark: evo-devo-neuroethology of electric communication in mormyrid fishes. *J. Neurogenet.* 27:106–129.
- Carlson, B. A., S. M. Hasan, M. Hollmann, D. B. Miller, L. J. Harmon, and M. E. Arnegard. 2011. Brain evolution triggers increased diversification of electric fishes. *Science* 332:583–586.
- Chapal, M., S. Mintzer, S. Brodsky, M. Carmi, and N. Barkai. 2019. Resolving noise-control conflict by gene duplication. *PLOS Biol.* 17.

- Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* 9:938–950. Nature Publishing Group.
- De Coster, W., S. D’Hert, D. T. Schultz, M. Cruts, and C. Van Broeckhoven. 2018. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669.
- Ding, B., D. W. Daugherty, M. Husemann, M. Chen, A. E. Howe, and P. D. Danley. 2014. Quantitative genetic analyses of male color pattern and female mate choice in a pair of cichlid fishes of Lake Malawi, East Africa. *PLOS One* 9:1–22.
- Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Gallant, J. R., M. E. Arnegard, J. P. Sullivan, B. A. Carlson, and C. D. Hopkins. 2011. Signal variation and its morphological correlates in *Paramormyrops kingsleyae* provide insight into the evolution of electrogenic signal diversity in mormyrid electric fish. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 197:799–817.
- Gallant, J. R., M. Losilla, C. Tomlinson, and W. C. Warren. 2017. The Genome and Adult Somatic Transcriptome of the Mormyrid Electric Fish *Paramormyrops kingsleyae*. *Genome Biol. Evol.* 9:3525–3530.
- Ganss, B., and W. Hoffmann. 2009. Calcium-induced conformational transition of trout ependymins monitored by tryptophan fluorescence. *Open Biochem. J.* 3:14–17.
- Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
- Gupta, R., and S. Brunak. 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symp. Biocomput.* 7:310–322.
- Hall, M. R., K. M. Kocot, K. W. Baughman, S. L. Fernandez-Valverde, M. E. A. Gauthier, W. L. Hatleberg, A. Krishnan, C. McDougall, C. A. Motti, E. Shoguchi, T. Wang, X. Xiang, M. Zhao, U. Bose, C. Shinzato, K. Hisata, M. Fujie, M. Kanda, S. F. Cummins, N. Satoh, S. M. Degnan, and B. M. Degnan. 2017. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* 544:231–234. Nature Publishing Group.
- Hoffmann, W. 1994. Ependymins and their potential role in neuroplasticity and regeneration: calcium-binding meningeal glycoproteins of the cerebrospinal fluid and extracellular matrix. *Int J Biochem* 26:607–619.
- Hopkins, C. D. 1981. On the diversity of electric signals in a community of mormyrid electric fish in West Africa. *Am. Zool.* 21:211–222.
- Hopkins, C. D. 1999. Signal evolution in electric communication. Pp. 461–491 in M. Hauser and M. Konishi, eds. *The design of animal communication*. MIT Press, Cambridge, MA.

Hoy, R. R., J. Hahn, and R. C. Paul. 1977. Hybrid cricket auditory behavior: Evidence for genetic coupling in animal communication. *Science* 195:82–84.

Jackson, D. J., C. McDougall, K. Green, F. Simpson, G. Wörheide, and B. M. Degnan. 2006. A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol.* 4:1–10.

Jensen, J. D., B. A. Payseur, W. Stephan, C. F. Aquadro, M. Lynch, D. Charlesworth, and B. Charlesworth. 2019. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution* 73:111–114.

Kern, A. D., and M. W. Hahn. 2018. The neutral theory in light of natural selection. *Mol. Biol. Evol.* 35:1366–1371.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–626.

King, J. L., and T. H. Jukes. 1969. Non-Darwinian evolution. *Science* 164:788–798.

Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.

Kosakovsky Pond, S. L., A. F. Y. Poon, R. Velazquez, S. Weaver, N. L. Hepler, B. Murrell, S. D. Shank, B. R. Magalis, D. Bouvier, A. Nekrutenko, S. Wisotsky, S. J. Spielman, S. D. W. Frost, and S. V. Muse. 2020. HyPhy 2.5 - A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* 37:295–299.

Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23:1891–1901.

Kosakovsky Pond, S. L., S. R. Wisotsky, A. Escalante, B. R. Magalis, and S. Weaver. 2021. Contrast-FEL-A Test for Differences in Selective Pressures at Individual Sites among Clades and Sets of Branches. *Mol. Biol. Evol.* 38:1184–1198.

Kramer, B. 1974. Electric organ discharge interaction during interspecific agonistic behaviour in freely swimming mormyrid fish. *J. Comp. Physiol.* 93:203–235. Springer-Verlag.

Kreitman, M. 1996. The neutral theory is dead. Long live the neutral theory. *BioEssays* 18:678–683.

Lavoué, S., M. E. Arnegard, J. P. Sullivan, and C. D. Hopkins. 2008. *Petrocephalus* of Odzala offer insights into evolutionary patterns of signal diversification in the Mormyridae, a family of weakly electrogenic fishes from Africa. *J. Physiol. Paris* 102:322–339.

Lavoué, S., J. P. Sullivan, and C. D. Hopkins. 2003. Phylogenetic utility of the first two introns of the S7 ribosomal protein gene in African electric fishes (Mormyroidea: Teleostei) and congruence with other molecular markers. *Biol. J. Linn. Soc.* 78:273–292.

Lefort, V., J. E. Longueville, and O. Gascuel. 2017. SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* 34:2422–2424.

- Li, W., and A. Godzik. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
- Lissmann, H. W., and K. E. Machin. 1958. The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J. Exp. Biol.* 35:451–486.
- Losilla, M., D. M. Luecke, and J. R. Gallant. 2020. The transcriptional correlates of divergent electric organ discharges in *Paramormyrops* electric fish. *BMC Evol. Biol.* 20:6.
- Löytynoja, A., and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U.S.A.* 102:10557–10562.
- Mason, N. A., K. J. Burns, J. A. Tobias, S. Claramunt, N. Seddon, and E. P. Derryberry. 2017. Song evolution, speciation, and vocal learning in passerine birds. *Evolution* 71:786–796.
- McDougall, C., M. J. Hammond, S. C. Dailey, I. M. L. Somorjai, S. F. Cummins, and B. M. Degnan. 2018. The evolution of ependymin-related proteins. *BMC Evol. Biol.* 18:182. *BioMed Central*.
- Mendelson, T. C., and K. L. Shaw. 2005. Rapid speciation in an arthropod. *Nature* 433:375. *Nature Publishing Group*.
- Möhres, F. P. 1957. Elektrische Entladungen im Dienste der Revierabgrenzung bei Fischen. *Naturwissenschaften* 44:431–432.
- Müller-Schmid, A., H. Rinder, F. Lottspeich, E.-M. Gertzen, and W. Hoffmann. 1992. Ependymins from the cerebrospinal fluid of salmonid fish: gene structure and molecular characterization. *Gene* 118:189–196.
- Murrell, B., S. Weaver, M. D. Smith, J. O. Wertheim, S. Murrell, A. Aylward, K. Eren, T. Pollner, D. P. Martin, D. M. Smith, K. Scheffler, and S. L. Kosakovsky Pond. 2015. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32:1365–1371.
- Murrell, B., J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. Kosakovsky Pond. 2012. Detecting individual sites subject to episodic diversifying selection. *PLOS Genet.* 8:e1002764.
- Nei, M., Y. Suzuki, and M. Nozawa. 2010. The neutral theory of molecular evolution in the genomic era. *Annu. Rev. Genomics Hum. Genet.* 11:265–289.
- Ohta, T. 1996. The neutral theory is dead. The current significance and standing of neutral and nearly neutral theories. *BioEssays* 18:673–677.
- Ortí, G., and A. Meyer. 1996. Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleost fishes. *Mol. Biol. Evol.* 13:556–573.
- Park, J. K., K. Y. Kim, Y. W. Sim, Y. I. Kim, J. K. Kim, C. Lee, J. Han, C. U. Kim, J. E. Lee, and S. Y. Park. 2019. Structures of three ependymin-related proteins suggest their function as a hydrophobic molecule binder. *IUCrJ* 6:729–739. *International Union of Crystallography*.

- Picq, S., J. Sperling, C. J. Cheng, B. A. Carlson, and J. R. Gallant. 2020. Genetic drift does not sufficiently explain patterns of electric signal variation among populations of the mormyrid electric fish *Paramormyrops kingsleyae*. *Evolution* 1–25.
- Pradel, G., M. Schachner, and R. Schmidt. 1999. Inhibition of memory consolidation by antibodies against cell adhesion molecules after active avoidance conditioning in zebrafish. *J. Neurobiol.* 39:197–206.
- Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4:1–8.
- Roy, A., A. Kucukural, and Y. Zhang. 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5:725–738.
- Schmidt, R., W. Brysch, S. Rother, and K. Schlingensiepen. 1995. Inhibition of Memory Consolidation After Active Avoidance Conditioning by Antisense Intervention with Ependymin Gene Expression. *J. Neurochem.* 65:1465–1471.
- Schwarz, H., A. Müller-schmid, and W. Hoffmann. 1993. Ultrastructural localization of ependymins in the endomeninx of the brain of the rainbow trout: possible association with collagen fibrils of the extracellular matrix. *Cell Tissue* 417–425.
- Shashoua, V. E. 1985. The role of brain extracellular proteins in neuroplasticity and learning. *Cell. Mol. Neurobiol.* 5:183–207.
- Shashoua, V. E., G. W. Hesse, and B. Milinazzo. 1990. Evidence for the in vivo polymerization of ependymin: a brain extracellular glycoprotein. *Brain Res.* 522:181–190.
- Shaw, K. L., and S. C. Lesnick. 2009. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl. Acad. Sci. U.S.A.* 106:9737–9742. National Academy of Sciences.
- Shen, W., S. Le, Y. Li, and F. Hu. 2016. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS One* 11:1–10.
- Smith, M. D., J. O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, and S. L. Kosakovsky Pond. 2015. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32:1342–1353.
- Staats, K. A., T. Wu, B. S. Gan, D. B. O’Gorman, and R. A. Ophoff. 2016. Dupuytren’s disease susceptibility gene, EPDR1, is involved in myofibroblast contractility. *J. Dermatol. Sci.* 83:131–137. Elsevier Ireland Ltd.
- Straka, H., J. Simmers, and B. P. Chagnaud. 2018. A New Perspective on Predictive Motor Signaling. *Curr. Biol.* 28:R232–R243. Elsevier Ltd.
- Suárez-Castillo, E. C., and J. E. García-Arrarás. 2007. Molecular evolution of the ependymin protein family: a necessary update. *BMC Evol. Biol.* 7:23.

- Sullivan, J. P., S. Lavoué, M. E. Arnegard, and C. D. Hopkins. 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* 58:825–41.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2002. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution* 56:597–616. Society for the Study of Evolution.
- Sullivan, J. P., S. Lavoué, and C. D. Hopkins. 2000. Molecular systematics of the African electric fishes (Mormyroidea: teleostei) and a model for the evolution of their electric organs. *J. Exp. Biol.* 203:665–683.
- Szabo, T. 1961. Les Organes Electriques des Mormyrides. Pp. 20–24 in C. Chagas and A. de Carvalho, eds. *Bioelectrogenesis*. Elsevier, New York.
- Taylor, J. S., and J. Raes. 2004. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* 38:615–643.
- von der Emde, G., M. Amey, J. Engelmann, S. Fetz, C. Folde, M. Hollmann, M. Metzen, and R. Pusch. 2008. Active electrolocation in *Gnathonemus petersii*: Behaviour, sensory performance, and receptor systems. *J. Physiol. Paris* 102:279–290. Elsevier Ltd.
- Wang, J., P. Youkharibache, D. Zhang, C. J. Lanczycki, R. C. Geer, T. Madej, L. Phan, M. Ward, S. Lu, G. H. Marchler, Y. Wang, S. H. Bryant, L. Y. Geer, and A. Marchler-Bauer. 2020. iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* 36:131–135.
- Weaver, S., S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, and S. L. Kosakovsky Pond. 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35:773–777.
- Wei, Y., Z. J. Xiong, J. Li, C. Zou, C. W. Cairo, J. S. Klassen, and G. G. Privé. 2019. Crystal structures of human lysosomal EPDR1 reveal homology with the superfamily of bacterial lipoprotein transporters. *Commun. Biol.* 2:1–13. Springer US.
- Wertheim, J. O., B. Murrell, M. D. Smith, S. L. Kosakovsky Pond, and K. Scheffler. 2015. RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32:820–832.
- Wiley, C., C. K. Ellison, and K. L. Shaw. 2012. Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. *Proc. Royal Soc. B* 279:1203–1209. The Royal Society.
- Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. 2015. The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* 12:7–8. Nature Publishing Group.
- Yang, J., and Y. Zhang. 2015. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* 43:W174–W181.
- Zakon, H. H., D. J. Zwickl, Y. Lu, and D. M. Hillis. 2008. Molecular evolution of communication signals in electric fish. *J. Exp. Biol.* 211:1814–1818.

Zhang, J. 2003. Evolution by gene duplication: An update. *Trends Ecol. Evol.* 18:292–298.

Zheng, F. X., X. Q. Sun, B. H. Fang, X. G. Hong, and J. X. Zhang. 2006. Comparative analysis of genes expressed in regenerating intestine and non-eviscerated intestine of *Apostichopus japonicus* Selenka (Aspidochirotida: Stichopodidae) and cloning of ependymin gene. *Hydrobiologia* 571:109–122.