Comparative Transcriptome Analyses of Key Components in the Convergent Evolution of Electric Organs in *Leucoraja erinacea*: An Exploratory Review

Chowdhury, F.H.[1](ID:20522336), Huynh, S. L.[2](ID:20583575), and Chan, D.[3](ID:20630060) (2017). University of Waterloo. BIOL 469 - Genomics. Instructor - Dr. Andrew C. Doxey.

1 - Project planner, data gathering. 2. Molecular interpretations and review. 3. Data analyst and programming.

Main Paper: Gallant, R. G., Traeger, L. L., Volkening, J. D., Moffett, H., Chen, H., Novina, C. D., ... Sussman, M. R. (2014). Genomic basis for the convergent evolution of electric organs. *Science,* 344(6191), 1522-1525.

1. Review and Background on the Paper by Gallant and Colleagues

1.1. Introduction to Electric Organs and Evolution

The study of electric organs (EO, EOs) in aquatic vertebrates is of significant interest to neuroethologists, scientists who use evolutionary and comparative approaches to study underlying mechanisms under the control of the nervous system (e.g. muscle contractions), as well as to evolutionary biologists. What makes the electric organs in fish so unique is not just its impressive function in generating electric currents outside the body, but rather its unexpected multiple occurrence in multiple lineages. The evolution of electric organs has been determined

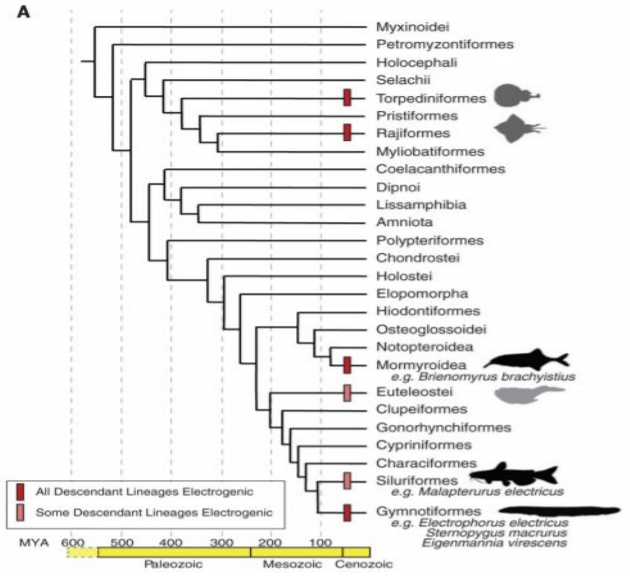

Figure 1. Phylogenetic Tree of Major Vertebrates Representing the Independent Events of Evolution of Electric Organs in Fishes

The following phylogenetic tree is a paraphyletic grouping of vertebrates that have been informally classified as the "fishes" which excludes tetrapods (i.e. birds, reptiles, amphibians, and mammals). It spans 25 Geological periods over ~ 500 million years (see scale bar in yellow). Reddish bars represent lineages that have evolved electric organs independently (6 in total). Black silhouettes represent lineages that were studied representing 3 independent evolutionary events of the appearance of electric organs. While grey silhouettes represent lineages that were not studied. The 3 lineages that were studied represent members from the major class of fishes, the Class: Actinopterygii (bony fish), which includes the Order: Gymnotiformes (knifefish, 3 specimens), and a closely related group, Order: Siluriformes (catfish, 1 specimen), and a more distant group, Order: Osteoglossiformes - Family: Mormyroidea (elephant fish, 1 specimen), totaling to 5 specimens representing 3 independent lineages portraying electric organs. Note that the study excluded the distant Class: Elasmobranchii (cartilaginous fish), which was represented by the Order: Rajiformes (skates) and the Order: Torpediniformes (electric rays). **SOURCE:** Main Paper

to have occurred in 6 lineages of aquatic vertebrates (i.e. fishes) all independently through convergent evolution. The idea is that convergent evolution is not as rare or by-chance as

previously predicted, and there are key principle mechanisms behind how it occurs. Figure 1 neatly explains and summarizes the lineages of fish that have evolved electric organs independently. Electric organs are specialized tissues composed of cells called electrocytes that are polar, innervated on one side by neurons, and stacked tightly together in a battery-like structure (Figure 2). They evolved as an adaptation mainly in fish for communication, navigation, and as a form of defense or predation. Electric organs are limited to aquatic vertebrates as they are externally discharged currents into the environment. The only suitable medium for such a feature is water, thus no terrestrial vertebrate has yet been described with such capabilities.

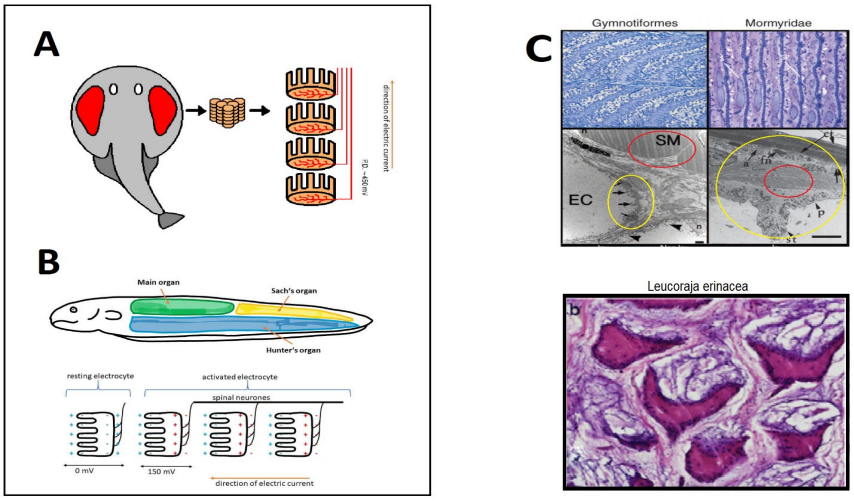

Figure 2. Basic Anatomy of Electric Organs & Structure of Electrocytes in Several Lineages of Fish (Vertebrates)
The following graphical is a basic portrayal of the function and morphology of electric organs found in various lineages. **A)** General model for electric organs in Order: Torpediniformes (electric rays). The EOs (in red) are composed of electrocytes (right) arranged in stacked like formation with invaginations facing upwards. It is important to note that Order: Torpediniformes (electric rays) and Order: Rajiformes (skates) evolved EOs independently, though this diagram is informally meant to represent the EOs of both lineages with respect to Class: Elasmobranchii (cartilaginous fish). It is key to know that the organism of interest for this research, *Leucoraja erinacea* (little skate) actually has its EOs located near the tail region, and not near the lateral head regions. **B)** General model for the electric organs in Order: Gymnotiformes (knifefishes) with emphasis on the electric eel. The electric eel has 3 EOs; the Main organ (green), Sach's organ (yellow), and Hunter's organ (blue). **C)** Histochemical analysis of the cellular morphology and features of electrocytes in Order: Gymnotiformes (knifefish) and Order: Osteoglossiformes - Family: Mormyroidea (elephant fish). In the middle left, an electrocyte is seen from a Gymnotiformes (yellow circle) which is absent of muscle-contraction complex called a sarcomere (see in red, striations of sarcomeres in muscle cell). In the middle right, an electrocyte from a Mormyroidea is seen (yellow) which appears to still have structures resembling sarcomeres, but are disorganized and non-functional. Bottom, electrocytes of Rajiformes (little skate; *Leucoraja erinacea*). Stacking, asymmetry, and lack of sarcomeres are seen. **SOURCE:** A (top left) - (Tanaka et al., 2016), B (bottom left) - (Soklic, 2017), C (right) - Main Paper + (Morson & Morrissey, 2007)

1.2. Rationale & Purpose of Studying Electric Organs in Fish

Previous work has found that electric organs (and their electrocytes) across all lineages appeared to have arisen from myogenic precursors (muscle cells). They are incredibly similar morphologically and functionally from lineage to lineage. Since the lineages in which they appear are far too diverged for them to have arisen through a common ancestor, it is further evidence that supports the idea that through independent convergent mechanisms acting on common starting point (i.e. myocytes in all lineages) was the reason behind how electrocytes in each lineage became extremely similar to each other.

By studying the similarities between the mechanisms and structures of independently evolving electric organs in many lineages of fish, Gallant and colleagues wish to identify a common molecular or genomic basis in all lineages in which analogous modification and mechanisms can act upon to contribute to development of electric organs and their functions (i.e. the analogous trait). The researchers believe such findings can aim to better explain the underlying principles of convergent evolution among distant lineages of not just electric organs in fish but for other complex adaptations in other species and lineages as well. For example, recent studies have found that spider silk in orbweavers may have evolved multiple times and is not ancestrally inherited from a common ancestor (Fernandez et al., 2014). In addition, the researcher believe that their findings can further advance the understanding of evolution of complex electrical signalling systems in humans such as the cardiac muscle conduction system.

1.3. Methods & Results: Searching for Electrocyte Specific Regulation of Genes

The overall methods from the paper revolved mainly around species belonging to the order Gymnotiformes (knifefish) as the basis for their model for evolution of electric organs with

emphasis on the electric eel (*Electrophorus electricus*). The idea was to construct a transcriptome comparison of several species belonging to the Gymnotiformes. Using the initial transcriptome data, the researchers aimed to find similar transcript patterns in 2 additional lineages of fish that have independently evolved electric organs, and belonging to the same class as the Gymnotiformes, which is the class Actinopterygii (bony fishes). Under this method, the researchers aimed to narrow down the common genetic components across all the lineages of interest that were targeted by analogous mechanisms in each lineage leading to the development of electric organs.

Next-Generation sequencing (NGS) of the electric eel was performed to construct its genome as a basis. Using informative gene prediction methods, a transcriptome for the electric eel (*Electrophorus electricus*) was created by analysing its electric organs (Figure 2B), as well as other non-electric organ related tissues such muscle tissue. The findings concluded that the electric eel genome consists of ~22, 000 protein coding genes.

The researchers employed variance filtering and k-mean clustering (k=12) to their results to detect tissue specific co-transcriptionally regulated genes. This narrowed down genes that were uniquely regulated in electric organs. This effectively narrowed down from 22, 000 genes to 9, 211 up-regulated genes and 1, 186 down-regulated genes (~10, 000 genes total).

The issue here is that the findings only represents regulated genes in the electric eel, and since its electric organs are analogous to electric organs of species outside the order Gymnotiformes, they cannot all be assumed as universal to electric organ development across all the other lineages. To solve this issue, a de novo transcriptome assembly was performed for 4 additional electric organ containing fish similarly as done in the eel. **(Summary below)**

Summary of all organisms used for the transcriptomes:

Class: Actinopterygii (bony fish) - 5 spp. representing 3 independent lineages

Order: Gymnotiformes (knifefish) - 3 species: ***Electrophorus electricus, Sternopygus macrurus, Eigenmannia virescens***

Order: Siluriformes *(*electric catfish*)* - 1 species: ***Malapterurus electricus* -** note that this is closely related group to Gymnotiformes (see Figure 1)

Order: Osteoglossiformes; Family: Mormyroidea (elephant fish) - 1 species: ***Brienomyrus brachyistius* -** note that this is also closely related to Gymnotiformes but less than the catfish (see Figure 1)

The researchers performed a reciprocal BLAST of eel transcriptome onto the additional 4 species transcriptomes. This narrowed the results down to ~100 commonly regulated genes.

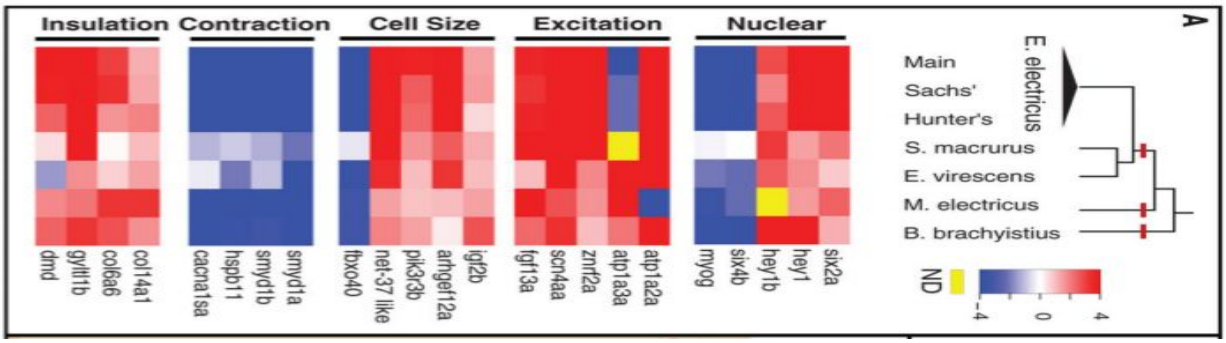

Figure 3. Key Genetic Markers from RNAseq Analyses for 3 Lineages of Fish that Independently Evolved Electric Organs

The following graphical is a heat-map for the significant up-regulated or down-regulated genes detected from RNAseq analysis of transcriptomes of electric organs from 5 species representing 3 lineages of independent EO evolution. Red bars on the tree represents independent evolutionary events of the electric organs. The level of up-regulation and down-regulation of each gene was calculated relative to the transcription levels from skeletal muscle cells from each lineage. Red boxed -> up-regulation. Blue boxed -> down-regulation. White boxed → no significant change. Yellow boxed → data deficient. Genes selected were based on 5 categories unique to electrocyte features: (i) nuclear transcription factors, (ii) genes that regulate cell excitation, (iii) genes that regulate cell size, (iv) genes involved in contraction and excitation contraction coupling, and (v) genes encoding proteins that surround individual electrocytes to provide the scaffold for insulation. Examples: (i) nuclear transcription factors: up-regulation of *hey1* gene - suppression of differentiation into muscle cells (ii) genes that regulate cell excitation: up-regulation of *scn4aa* - voltage dependant ion channels in (iii) genes that regulate cell size: down-regulation of *fbxo40* gene - promoting muscle differentiation and growth and suppressing electrocyte related growth (iv) genes involved in contraction and excitation contraction coupling: down-regulation of *cacna1sa* gene - calcium channel involved in excitation-contraction coupling (v) genes for insulation: up-regulation of *col14a1*- collagen related genes involved in insulation and scaffolding of electrocyte cells

SOURCE: Main Paper

In order to further narrow down the genes of interest, and RNAseq analyses was performed (Figure 3.) to analyse patterns of up down regulated genes in electrocytes in comparison to other muscle derived cells (e.g. cardiac, skeletal) across all lineages. Since the goal of the experiment was to determine genes unique to the development and function of electrocytes, the genes of interest were narrowed down based on criteria specific to electrocyte features relative to muscle cells (recall that they are muscle derived). The filtering was based on 5 functional criteria (see Figure 3

for list). This narrowed down the findings to ~25 key regulated genetic targets closely linked to electric organ development across all lineages.

1.4. Significance

The mechanisms of convergent evolution is based on repeated targeting of common starting developmental pathways that ultimately lead to same functional adaptation. It is now believed that convergent evolution of complex traits is not entirely random, and independent systems are more similar than previously believed to be at the molecular level.

2. Objectives of Comparative Analysis - Exploring Other Lineages of Electric Fish

Despite the findings from Figure 1, where the class Elasmobranchii (cartilaginous fish), in which 2 specific lineages that also independently evolved electric organs which includes the lineages from the order Rajiformes (skates) and the order Torpediniformes (electric rays), were not included in their analysis. This means they have not extensively proven that their findings can be applied universally to all electric fish lineages, and is the basis of this analysis and review.

For this analysis, the idea is to include another lineage from the Elasmobranchii from the order Rajiformes (skates) in addition to paper by comparing the transcriptome set of the Little Skate (*Leucoraja erinacea*) to the transcriptome sets prepared by the authors. This species was chosen for the analysis since it is a commonly used model organism in biomedical research (Wyffels et al., 2014). The goal is to identify key genetic components that were characterized in the paper that were shown to have been found to be highly pronounced in terms of change in expression levels in all lineages of electric organ containing fish. Since RNAseq cannot be performed, this analysis only acts to validate that the Elasmobranchii contain common genes in electrocyte

development, and is meant to pave the way for follow up analysis similar to the RNAseq performed in the paper.

3. Comparative Analysis - Methods & Findings

Three main data samples were acquired: hand-picked genes from *Danio rerio,* which were sourced from the NCBI gene database; An *E. electricus* transcriptome sequenced and assembled by Gallant et al. (2014); and a transcriptome of *L. erinacea* sequenced and assembled by King et al. (2011).

Gallant and colleagues (2014) found key genes in the physiology of electric organs. Elementary analysis of RNA-Seq data was performed in order to test Gallant's claims on species of order Rajiformes (of which *L. erinacea* belongs). As of present, *L. erinacea* as an organism is not well annotated genomically. A good proxy to use for annotation via BLAST would be *E. electricus*. However, *E. electricus* itself was not annotated. *D. rerio*, as a well studied organism, was used extensively in annotating.

In this analysis, the use of contigs is essentially an approximation to a gene. Local BLASTN version 2.7.1+ was used to perform reciprocal BLAST searches on *E. electricus* and *L. erinacea* databases to determine orthologous contigs. Parameters were set by the preset DC-Megablast. For future reference, BLAST without parameters mentioned should be assumed to have been done via DC-Megablast.

Genes of interest from Gallant and colleagues (2014) include: *scn4aa, six2a, hey1, hey1b, six4b, myog, atp1a2a, atp1a3a, znrf2a, fgf13a, igf13a, igf2b, arhgef12a, pik3r3b, net-37, fbxo40, smyd1a, smyd1b, hspb11, cacna1sa, col14a1, col6a6, gyltl1b, dmd*.

Of the above genes found annotated in *D. rerio*, which have significant hits in the *E. electricus* transcriptome, and appear orthologous to contigs in *L. erinacea*, the list significantly falls to the following genes: *six2a, ATPase, scn4aa, pik3r3b, gylbtl1b, cacna1sa, hspb11*. Using statistical methods, it was determined whether these genes are over or under expressed by processes other than chance.

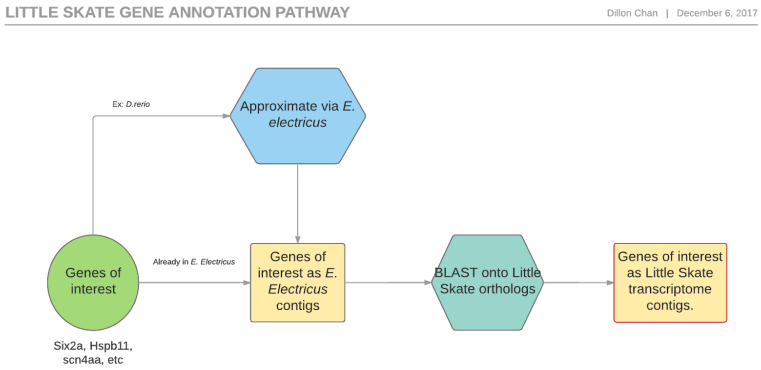

Figure 4. Annotation Pathway
Genes of interest included *scn4aa, six2a, hey1, hey1b, six4b, myog, atp1a2a, atp1a3a, znrf2a, fgf13a, igf13a, igf2b, arhgef12a, pik3r3b, net-37, fbxo40, smyd1a, smyd1b, hspb11, cacna1sa, col14a1, col6a6, gyltl1b, dmd*. Only *scn4aa* was found annotated in *E. electricus*. When BLASTed onto *L. erinacea* contigs, there were highly significant hits (E-Value < 1e-21) with high identity (> 70%). Genes of interest as *L. erinacea* contigs have significant (E-Value < 1e-15) hits.

Data from King and colleagues (2011) also included RNA-seq data. In particular, average coverage per contig. As an approximation to true expression data, the average coverage per contig was used as a indication of expression levels. Based on the work done by Gallant et al. (2014), it was predicted that *six2a, ATPase, scn4aa, pik3r3b, gylbtl1b* would be up-regulated, and *cacna1sa, hspb11* would be down-regulated.

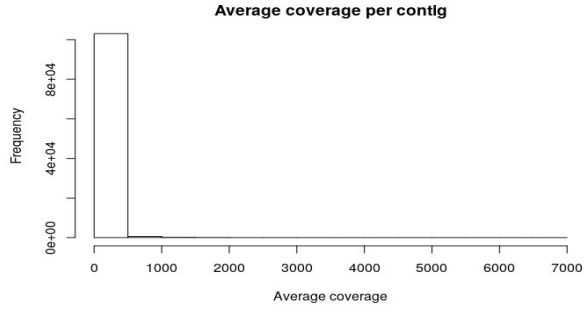

Figure 5. Raw Distribution of Contig Coverage
By far, the majority of contigs had average coverage in the [0,30] interval. Frequency for bins with more than 1000 average coverage were not considered outlier data because the frequency distribution decreases 'smoothly' although suddenly after coverage > 500. This introduces considerable skew (e.g., the mean coverage is 37.2 but the 3rd quartile is at 27.13) which is handled with a suitable power transformation.

A power transformation ($\alpha = -0.3144$) was applied to the average coverage distribution over all contigs. The power transformation is defined as:

$$T_\alpha(y) = \frac{y^\alpha - 1}{\alpha}$$

The coefficient α is chosen as the value that minimizes the median skewness of the distribution. The distribution is then normalized and assumed to be Gaussian. Locating the genes of interest on the unskewed, normalized distribution suggests that the there is possibility for some genes to be up-regulated or down-regulated by chance, i.e., insignificant results.

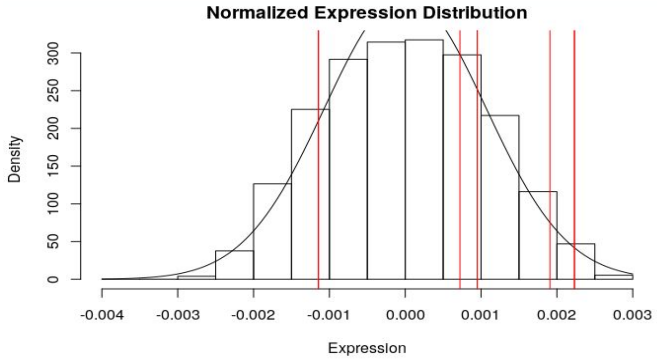

Figure 6. Normalized Coverage Distribution with Zero Median Skewness.
The distribution was assumed to be Gaussian for the purposes of this project, but further research should include determining with more certainty the correctness of this assertion. For example, *t* distributions may also be an appropriate model. Location of genes are marked with a vertical line.

4. Discussion and Future of Electric Fish Genomics

The goal of the project was to uncover the evolutionary relationship regarding the development of electric organs between the orders Rajiformes and Gymnotiformes by using Gallant's original methods and applying it to the *L. erinacea* transcriptome. Contrary to initial predictions, the results acquired from the crude gene annotation pipeline, which was designed to compare *E. electricus* genes to *L. erinacea* genes, were not completely representative of convergent evolution. Since the *L. erinacea* transcriptome was primarily sourced from embryonic tissue samples , which are not as indicative of gene expression as specialized adult tissues, the mere

presence of hits on target genes could not provide the conclusive proof needed (Wyffels et al., 2014). Nevertheless, it did provide some reasonable insight into the general up-regulation and down-regulation of certain genes of interest in *L. erinacea* when compared to *E. electricus*. Further examination of the evolutionary link between the orders using comparative analysis tools with adult *L. erinacea* tissue samples is needed, leaving great potential for future research on the subject.

The target genes, *six2a, ATPase, scn4aa, pik3r3b, gylbtl1b*, *cacna1sa,* and *hspb11* were all observed to be hits with varying accuracies. The hits were generally above ~70% identities, proving to be similar for the most part. The ~30% discrepancy can be expected due to the distant genetic gap between the lineages being examined: Rajiformes and Gymnotiformes; the conservation of sequence would be different due to genetic and physiological differences. However, this still does not prove the presence of convergent evolution, which was the aim of the project.

While the presence of *six2a, ATPase, scn4aa, pik3r3b,* and *gylbtl1b* was confirmed per expectations, the presence of *cacna1sa* and *hspb11* did not meet predictions (Figure 3). They were still present in the *L. erinacea* transcriptome despite being expected to be down-regulated. This can be explained by the fact that embryonic tissue was used and the up-regulation and down-regulation of genes are not as prominent as they would be in adult tissues. Furthermore, the target genes did not show any sign of considerable up- or down-regulation when compared to other prominent electric organ genes, such as *myog, atp1a2a,* and *fgf13a*. With such great variances and inaccuracies in expected and actual outcomes, the results were considered to be inconclusive and not representative of convergent evolution.

The key limitations faced throughout the comparative analysis process were a lack of representative and properly annotated transcriptomes and the lack of time in advancing the project to the next stage of analysis. Regarding the genomic data *L. erinacea*, there was limited high-quality data to work with (Wyffels et al., 2014). All the sequences uploaded to databases, such as NCBI, belonged to either adult or embryonic samples derived from multiple tissues (Wyffels et al., 2014). The adult sample would be the first reasonable choice to work with but they lacked the crucial electric organ tissues in the mix of source tissues, rendering it unusable for the project (Wyffels et al., 2014). On the other hand, the embryonic samples, sourced from stage 19, 20, 25 and 28 embryos, did contain the electric organ tissues but were not a concrete representation of actual gene expression in adult *L. erinacea* (Wyffels et al., 2014). Since the *E. electricus* transcriptome was sourced from an adult, the analysis calls for the same filter to be placed on any *L. erinacea* samples used.

Even though the embryonic *L. erinacea* samples provided a general overview of the presence and absence of the key genes being scouted for, all scientific studies need to uphold a standard of credibility by maintaining the consistency and relevance of the results. As such, the scope of this project was not possible to follow, unless an adult *L. erinacea* sample with the electric organ tissues was acquired, or a pivot in the species or order studied as considered.

Since the results from the comparative analysis of *L. erinacea* and *E. electricus* were not as hypothesized, a change in target lineage was proposed. The order Torpediniformes, one of the three lineages that Gallant did not cover in his original study (Figure 1), was picked as the next ideal prospect for examination due to a recent breakthrough in genome sequencing with a representative species, *Tetronarce californica* (Stavrianakou et al., 2017). However, only the

individual reads of the *T. californica* transcriptome are currently available (Stavrianakou et al., 2017). As a result, a *de novo* assembly of its transcriptome would have to be carried out for the comparative analysis with *E. electricus*. This process would enable a search for highly conserved regions in an attempt to identify genetic differences in the electric organs of both species. Unfortunately, due to a lack of time in completing the project and the need for information recycling due to the pivot, the milestone was not reached and still remains a goal for the future.

The accumulation of sequence data is of the highest importance in regards to developing the budding field of electric fish genomics (Pitchers at al., 2016). The use of genomic tools would provide researchers with sequence data that would then allow for further refinement of the said tools, creating a positive feedback cycle of progress. It would also allow for relevant protocols to be created and maintained, and promote future research projects on electric fish.

Studies that examine gene and protein expression pathways can help determine the effects of loss of function (Pitchers at al., 2016). They can also provided valuable information on gene regulatory networks and highlight the crucial biological processes that take place on the transcriptional and translational level in electric organs (Pitchers at al., 2016). In addition, researchers can use RNAseq profiling to detect differentially expressed genes and pathways, which would also aid in discovering alternative transcripts. All these strategies and protocols need to be developed and fully realized for electric fish genomics to truly scale throughout both academia and industry for future applications (Pitchers at al., 2016).

The potential applications of electric fish as new model organisms are vast in scale and have the ability to provide many insights into the scientific fields, including evolutionary biology, bioinformatics, biotechnology, biophysics, and medicine. The research carried out by Gallant and his

team, along with other prominent researchers, have propelled electric fish into the spotlight of genomics, but much work is still required to fully develop the molecular tools that would expedite future discoveries (Pitchers et al., 2016). Along with the delivery of scientific advances, these tools and protocols will play an integral part in answering some of the most challenging questions surrounding evolution and synteny; more specifically, it would help further understanding of convergent evolution. With the widespread adoption of affordable, automated technologies and the accumulation and sequencing of genomic data, the electric fish community can become a global leader in next-generation biology and comparative genomics.

5. Acknowledgements

The authors wish to thank J. Gallant and his colleagues for their unparalleled contributions to electric fish genomics. Additional acknowledgement goes to the organizer of the BIOL 469 course, A. Doxey, for his teachings and support throughout the duration of the project. Lastly, gratitude goes to the UW Computer Science Club for graciously allowing use of their machines.

6. References

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research, 25*(17), 3389-3402. doi:10.1093/nar/25.17.3389

Fernandez, R., Hormiga, G., and Giribet, G. (2014). Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers. *Current Biology*, 24, 1-6.

Gallant, J. R., Traeger, L. L., Volkening, J. D., Moffett, H., Chen, P., Novina, C. D., . . . and Sussman, M. R. (2014). Genomic basis for the convergent evolution of electric organs. *Science, 344*(6191), 1522-1525. doi:10.1126/science.1254432

King, B. L., Gillis, J. A., Carlisle, H. R., and Dahn, R. D. (2011). A natural deletion of the HoxC cluster in elasmobranch fishes. *Science*,334(6062), 1517.

Morson, J. M., and Morrissey, J. F. (2007). Morphological variation in the electric organ of the little skate (*Leucoraja erinacea*) and its possible role in communication during courtship. *Environmental Biology of Fishes*, Vol. 80, Issue 2-3, 267-275.

Pitchers, W. R., Constantinou, S. J., Losilla, M., & Gallant, J. R. (2016). Electric fish genomics: Progress, prospects, and new tools for neuroethology. Journal of Physiology Paris, 110(3), 259–272. doi: 10.1016/j.jphysparis.2016.10.003

Soklic, A. (2017). Biomimetic potential of electric eels. Retrieved from http://www.blogionik.org/biomimetic-potential-electric-eels/.

Stavrianakou, M., Perez, R., Wu, C., Sachs, M. S., Aramayo, R., & Harlow, M. (2017). Draft de novo transcriptome assembly and proteome characterization of the electric lobe of Tetronarce californica: a molecular tool for the study of cholinergic neurotransmission in the electric organ. BMC Genomics, 18(1), 611. doi: 10.1186/s12864-017-3890-4

Tanaka, Y., Funano, S., Nishizawa, Y., Kamamichi, N. Nishinaka, M., and Kitamori, T. (2016). An electric generator using living Torpedo electric organs controlled by fluid pressure-based alternative nervous system. *Scientific Reports*, 6.

Wyffels, J., King, B. L., Vincent, J., Chen, C., Wu, C. H., and Polson, S. W. (2014). SkateBase, and elasmobranch genome project and collection of molecular resources for chondrichthyan fishes, *F1000Research, 3, 191.*