



**Evolution and ontogeny of electric organ discharge in African weakly electric fish genus *Campylomormyrus*: a genomic and transcriptomic perspective**



Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

in der Wissenschaftsdisziplin Evolutionsbiologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

der Universität Potsdam

von

**Feng Cheng**

Potsdam, September 2023

This work is protected by copyright and/or related rights. You are free to use this work in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s).

<https://rightsstatements.org/page/InC/1.0/?language=en>

## **Gutachter**

**Prof. Dr. Ralph Tiedemann**

Universität Potsdam

**Prof. Dr. Jason Gallant**

Michigan State University

**Prof. Dr. Walter Salzburger**

Universität Basel

Published online on the

Publication Server of the University of Potsdam:

<https://doi.org/10.25932/publishup-63017>

<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-630172>

## **Erklärung der Urheberschaft**

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt.

---

Ort, Datum

Feng Cheng



## Abstract

The African weakly electric fishes (Mormyridae) exhibit a remarkable adaptive radiation possibly due to their species-specific electric organ discharges (EODs). It is produced by a muscle-derived electric organ that is located in the caudal peduncle. Divergence in EODs acts as a pre-zygotic isolation mechanism to drive species radiations. However, the mechanism behind the EOD diversification are only partially understood.

The aim of this study is to explore the genetic basis of EOD diversification from the gene expression level across *Campylomormyrus* species/hybrids and ontogeny. I firstly produced a high quality genome of the species *C. compressirostris* as a valuable resource to understand the electric fish evolution.

The next study compared the gene expression pattern between electric organs and skeletal muscles in *Campylomormyrus* species/hybrids with different types of EOD duration. I identified several candidate genes with an electric organ-specific expression, e.g. *KCNA7a*, *KLF5*, *KCNJ2*, *SCN4aa*, *NDRG3*, *MEF2*. The overall genes expression pattern exhibited a significant association with EOD duration in all analyzed species/hybrids. The expression of several candidate genes, e.g. *KCNJ2*, *KLF5*, *KCNK6* and *KCNQ5*, possibly contribute to the regulation of EOD duration in *Campylomormyrus* due to their increasing or decreasing expression. Several potassium channel genes showed differential expression during ontogeny in species and hybrid with EOD alteration, e.g. *KCNJ2*.

I next explored allele specific expression of intragenus hybrids by crossing the duration EOD species *C. compressirostris* with the medium duration EOD species *C. tshokwe* and the elongated duration EOD species *C. rhynchophorus*. The hybrids exhibited global expression dominance of

the *C. compressirostris* allele in the adult skeletal muscle and electric organ, as well as in the juvenile electric organ. Only the gene *KCNJ2* showed dominant expression of the allele from *C. rhynchophorus*, and this was increasingly dominant during ontogeny. It hence supported our hypothesis that *KCNJ2* is a key gene of regulating EOD duration. Our results help us to understand, from a genetic perspective, how gene expression effect the EOD diversification in the African weakly electric fish.

## Zusammenfassung

Die Mormyridae, eine Familie afrikanischer schwach elektrischer Süßwasserfische, zeigen eine außergewöhnliche adaptive Radiation. Eine Erklärung für die Diversifizierung dieser Gruppe stellen die artspezifischen elektrischen Organentladungen (EODs) dar. Diese werden von einem elektrischen Organ muskulären Ursprungs im Ansatz der Schwanzflosse erzeugt. Die verschiedenen EODs könnten als präzygotischer Isolationsmechanismus für die Radiation verantwortlich sein. Dennoch ist der Mechanismus hinter der EOD-Diversifizierung bisher nicht vollständig geklärt.

Ziel dieser Studie ist es, die genetische Grundlage der EOD-Diversifizierung auf der Ebene der Genexpression bei verschiedenen *Campylomormyrus*-Arten bzw. -Hybriden und während der Ontogenese zu ermitteln. Zunächst wurde erstmals das Genom der Art *C. compressirostris* in hoher Qualität sequenziert. Dies bildet eine bedeutende Grundlage für das Verständnis der Evolution der elektrischen Fische.

In der zweiten Studie wurden Genexpressionsmuster von elektrischen Organen und Skelettmuskeln bei *Campylomormyrus*-Arten bzw. -Hybriden mit unterschiedlicher EOD-Dauer verglichen. Dabei konnten mehrere Kandidatengene identifiziert werden, die potentiell Elektroorgan-spezifisch exprimiert sind, i.a. *KCNA7a*, *KLF5*, *KCNJ2*, *SCN4aa*, *NDRG3*, *MEF2*. Bei allen untersuchten Arten/Hybriden wies das Genexpressionsmuster einen signifikanten Zusammenhang mit der EOD-Dauer auf. Die Expression mehrerer Kandidatengene, wie beispielsweise *KCNJ2*, *KLF5*, *KCNK6* und *KCNQ5*, trägt möglicherweise zur Regulierung der EOD-Dauer bei *Campylomormyrus* bei. Bei Arten und Hybriden mit EOD-Unterschieden zeigten Kaliumkanal-Gene wie *KCNJ2* eine unterschiedliche Expression während der Ontogenese.

Zudem wurde die Allel-spezifische Expression bei Intragenus-Hybriden unter Verwendung der Arten *C. compressirostris*, *C. tshokwe* und *C. rhynchophorus*, die jeweils eine kurze, intermediäre bzw. lange EOD-Dauer aufweisen, untersucht. Die Hybriden wiesen eine generell dominante Expression der Allele von *C. compressirostris* in der adulten Skelettmuskulatur und im elektrischen Organ sowie im juvenilen elektrischen Organ auf. Einzig im Gen *KCNJ2* dominierte das Allel von *C. rhynchophorus*, mit zunehmender Dominanz mit fortschreitender Ontogenese. Dies stützt unsere Hypothese einer Beteiligung des *KCNJ2*-Gens an der Regulation der EOD-Dauer. Unsere Ergebnisse stellen einen wesentlichen Beitrag zum Verständnis des Einflusses der Genexpression auf die EOD-Diversifizierung bei afrikanischen schwach elektrischen Fischen dar.



# Contents

<b>Abstract</b> .....	- 1 -
<b>Zusammenfassung</b> .....	- 3 -
<b>1 Introduction</b> .....	- 6 -
<b>1.1 Whole genome duplication in teleost fish</b> .....	- 6 -
<b>1.2 Convergent evolution in electric fish</b> .....	- 6 -
<b>1.3 Adaptive radiation and electric organ discharges in mormyrids</b> .....	- 8 -
<b>1.4 Electric organ discharge in <i>Campylomormyrus</i></b> .....	- 9 -
<b>1.5 Intragenus hybridization in <i>Campylomormyrus</i></b> .....	- 12 -
<b>1.6 Aims of this study</b> .....	- 13 -
<b>2 Summary of articles</b> .....	- 16 -
<b>2.1 Summary of Article I</b> .....	- 16 -
<b>2.2 Summary of Article II</b> .....	- 17 -
<b>2.3 Summary of Article III</b> .....	- 18 -
<b>3 Article I (published)</b> .....	- 21 -
<b>4 Article II</b> .....	- 53 -
<b>5 Article III</b> .....	- 121 -
<b>6 General discussion</b> .....	- 162 -
<b>6.1 Genomics of <i>Campylomormyrus</i></b> .....	- 162 -
<b>6.2 What makes the EO?</b> .....	- 163 -
<b>6.3 EOD duration diversification among <i>Campylomormyrus</i> species</b> .....	- 164 -
<b>6.4 EOD development during ontogeny</b> .....	- 166 -
<b>6.5 EOD diversification in <i>Campylomormyrus</i> hybrids</b> .....	- 167 -
<b>7 Conclusion</b> .....	- 169 -
<b>8 Reference</b> .....	- 171 -
<b>9 Acknowledgements</b> .....	- 176 -

# **1 Introduction**

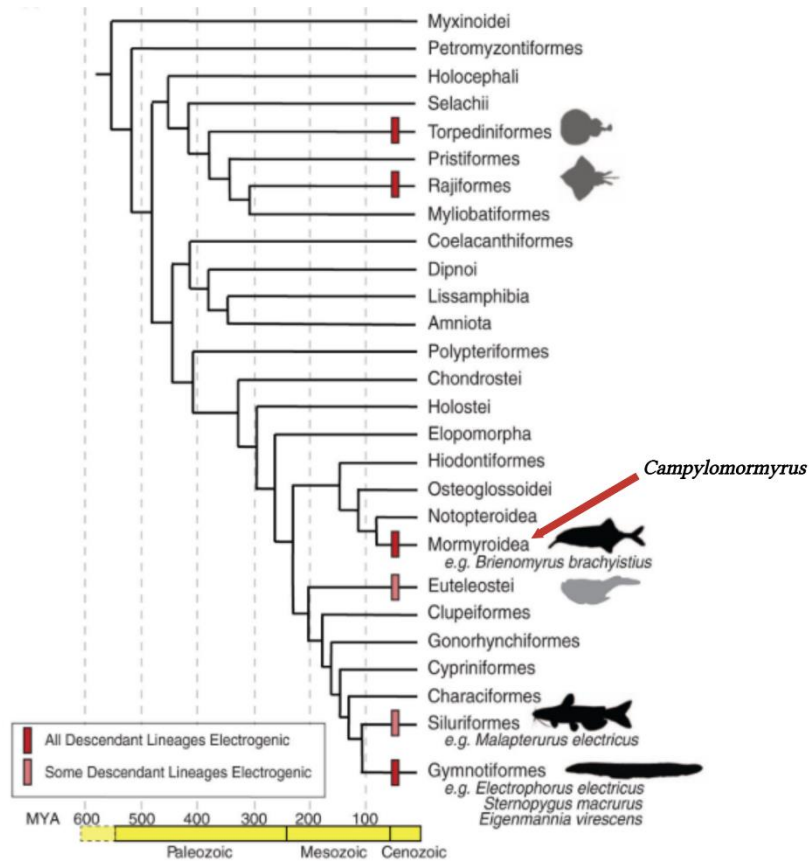
## **1.1 Whole genome duplication in teleost fish**

Whole genome duplication (WGD) events have played a crucial role in regulating the evolutionary history of many lineages (Hooper and Berg 2003; Jatllon et al. 2004; Volf 2005). Teleost fish, which constitute more than half of the extant vertebrates, contain a magnificent level of biodiversity (Volf 2005). There is now substantial evidence that a teleost-specific WGD has taken place during the early evolution of this lineage (Jatllon et al. 2004; Kasahara et al. 2007; Glasauer and Neuhauss 2014). WGD is typically followed by massive gene loss, genomic rearrangements and most importantly, genes retained in duplicate (Jatllon et al. 2004; Kasahara et al. 2007; Gundappa et al. 2022). WGD has long been recognized as a contributor to the evolution of genes with new functions, e.g. subfunctionalization (two copies split the initial function) and neofunctionalization (one copy can generate a novel function) (Hooper and Berg 2003). Although we lack evidence for the direct connection between the teleost-specific WGD and the magnificent biodiversity in teleost fish, differential neofunctionalization or subfunctionalization in duplicated genes may have been involved in the generation of fish variability, and consequently resulted in the successful radiation of teleost fishes (Hooper and Berg 2003).

## **1.2 Convergent evolution in electric fish**

The electric fish have independently evolved the ability to generate electric fields in at least six clades in teleost and elasmobranch fish: torpedinoids, rajoids, gymnotiforms, siluriforms, uranoscopids and mormyroids (Figure 1.1; Stern 2013; Gallant et al. 2014; Carlson et al. 2019). Electric fishes have the special ability to produce and perceive electric signals, the electric organ discharge (EOD), which is used for object sensing and social communication (weak EOD, e.g.

gymnotiforms and mormyroids) or for stunning prey and warding off predators (strong EOD, e.g. torpenid rays, electric eel; Crampton 2019).



**Figure 1.1** Phylogenetic tree of teleost fishes and major groups of electric fishes (Cited from Gallant *et al.* 2014).

The EOD, as signals of the nervous system, is in the realm of electricity (Bass 1986). Therefore, it can be easily comprehensible by the biophysics of ion currents. Available studies had identified several duplicated sodium and potassium channel gene copies, which are convergently expressed in electric fish clades. A striking example is the voltage-gated sodium channel gene *SCN4ab* which is still expressed in the skeletal muscle, while its paralog *SCN4aa* has radically shifted the expression to the muscle-derived electric organ (Gallant *et al.* 2014; Wang and Yang 2021). The shifted expression possibly promoted the evolution of a new function or even the electric organ itself (Thompson *et al.* 2014).

### **1.3 Adaptive radiation and electric organ discharges in mormyrids**

Mormyrids is a clade of freshwater weakly electric fish endemic to African riverine and partially lacustrine systems (Tiedemann et al. 2010). It comprises the superfamily Mormyroidea along with the family Gymnarchidae. There are more than 180 described species in at least 20 genera in mormyrids, which account for nearly 90% of the osteoglossomorph fishes (Feulner et al. 2008).

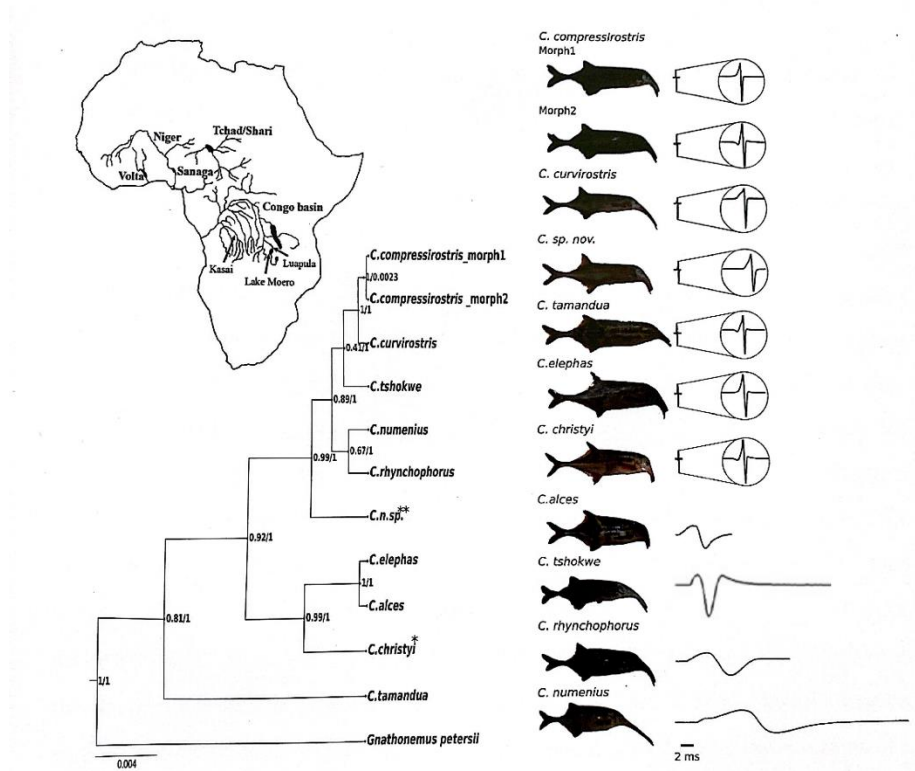
Adaptive radiation is generally the consequences of divergent natural selection along with reproductive isolation, which interrupts gene flow across diverging lineages. Frequently, reproductive isolation is caused by geographical distance called allopatry. However, in parapatry or sympatry, adaptive radiations associated with ecological niches diversification and are often leading to significant phenotypic and ecological diversification, necessitate a reproductive isolation mechanism. Previous studies suggest that a divergent EOD used for species recognition and mate choice comprises such a mechanism in mormyrids (Arnegard et al. 2005; Carlson et al. 2011).

EOD is produced from a myogenic electric organ that is located in the caudal peduncle in the adult mormyrids (Bennett 1970; Bass 1986). It is composed of specialized electrocytes that are longitudinally stacked (Bennett 1971). The sum of the action potential, which is propagated by each electrocyte, is the externally measurable EOD (Bennett 1970; Bennett 1971; Bass 1986).

EODs are very diverse both across and within genera in mormyrids. The EOD in mormyrids has the basic function of electrolocation, and is additionally used for species communication and mate recognition (Feulner et al. 2009a; Feulner et al. 2009b; Nagel, Kirschbaum, Hofmann, et al. 2018; Nagel, Kirschbaum, Engelmann, et al. 2018). It can serve as pre-zygotic isolation mechanism and lead to environmental niches variation, or triggers speciation by sexual selection. EODs in the

genus *Paramormyrops* show sexual dimorphism, which might indicate sexual selection (Arnegard et al. 2010). In the genus *Campylomormyrus* (Fig. 1), the EODs probably promote ecological speciation, while no sexual dimorphism was found in *Campylomormyrus* so far.

#### 1.4 Electric organ discharge in *Campylomormyrus*



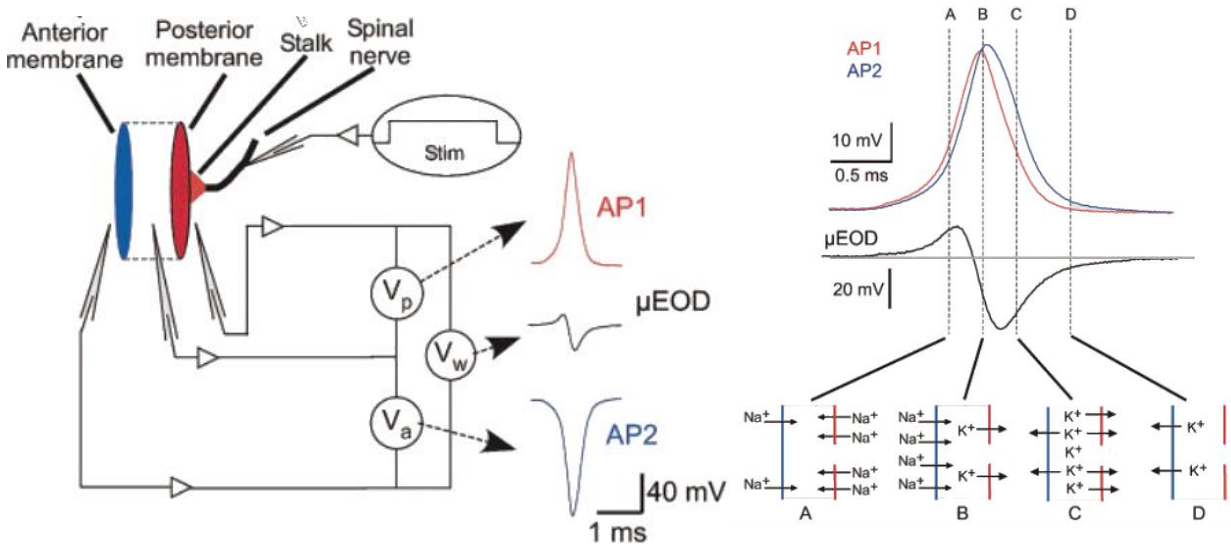
**Figure 1.2** Phylogenetic tree and electric organ discharges of *Campylomormyrus* species (Modified from Lamanna et al. 2016).

*Campylomormyrus* species are mostly distributed in Congo River system and its tributary streams (Figure 1.2). They are characterized by prominent, tubular and elongated snouts. There are 15 described species in this genus based on the morphometric measurements and EOD waveform (Lamanna et al. 2016).

Comparisons of EODs in *Campylomormyrus* revealed species-specific diversity in the shape and waveform (Figure 1.2). The EOD comprises the sum of action potentials that are produced by each

stalk and face of an electrocyte (Bennett 1970; Bass 1986). Therefore, the EOD types with different polarity and number of phases result from the geometrical diversity of the electrocytes (Bass 1986). EOD duration can vary 100-fold across species, e.g. *C. compressirostris* possess a short EOD around 0.4 ms while *C. rhynchophorus* and *C. numenius* have magnificently elongated EOD over 40 ms. In the adult *Campylomormyrus*, short EODs occurs in the most basal clade (*C. tamandua*). It is considered as a plesiomorphic feature, while the long duration (including medium duration) EOD is an apomorphic (derived) feature (Kirschbaum et al. 2016). One explanation for the elongation of the EOD is associated with an increased electrocyte surface, which is assumed to increase the membrane capacitance (Bennett 1970). In two species with very elongated EOD, *C. rhynchophorus* and *C. numenius*, a special structure, so called papillae, apparently contribute to their long EOD (Paul et al. 2015; Kirschbaum et al. 2016; Nguyen et al. 2020; Korniienko et al. 2021). These papillae are surface specializations of the uninnervated anterior face of the electrocyte (Korniienko et al. 2021). Despite the geometric differences in electrocytes among some species, the mechanism of waveform diversification is still poorly understood, since species with very diverged EOD waveform can possess a similar electric organ anatomy (Paul et al. 2015), e.g. *C. compressirostris* and *C. tshokwe* (around 5 ms). Modulated ion currents (especially potassium and sodium; Figure 1.3) potentially contribute to the EOD waveform difference, since the EOD is the sum of action potentials from all electrocytes (Nagel et al. 2017). This has been preliminary testified in a voltage-gated potassium channel gene *KCNA7a* that might contribute to the EOD duration in *Brienomyrus* and *Gymnarchus* (Swapna et al. 2018). In addition, a gene expression comparison in the electric organ between *C. compressirostris* (short duration EOD) and *C. tshokwe* (medium duration EOD) showed an upregulation in all *KCNA* genes in *C. tshokwe*, suggesting this

voltage-gated potassium channels to be potentially involved in EOD signal divergence in *Campylomormyrus* (Nagel et al. 2017).



**Figure 1.3** EOD of a single electrocyte from electric fish. (Cited from **Stoddard and Markham 2008**).

EOD variation in *Campylomormyrus* is not only exhibited across species, but also in different life stages (Nagel et al. 2017; Nguyen et al. 2020). A short EOD has been found in the juvenile stage of all species investigated so far. *C. compressirostris*, *C. tamandua*, and *C. tshokwe* start with juvenile EODs of identical shape and waveform during ontogeny, only *C. rhynchophorus* has same shape but twice as long as other three species (Nguyen et al. 2017). The EOD of *C. compressirostris* shows consistence over the whole length of the ontogeny. However, in the other studied species, the juvenile EODs continuously change during development until they reach the adult EOD (Nguyen et al. 2017; Nguyen et al. 2020). So far, we have only limited knowledge of how the EOD varies among species and during ontogeny. More detailed studies are therefore

needed to understand the phenotypic diversification among *Campylomormyrus* and its contribution to the observed radiations.






















### **1.5 Intragenus hybridization in *Campylomormyrus***

Successful hybridization had been performed using species with different EOD shape and waveform (Kirschbaum et al. 2016). This further contributes to our understanding of the EOD variation among species and ontogenic stages (Figure 1.4).

In the different cross combination of species with short EOD and longer EOD, the hybrids consistently show long EODs, even at an early life stage. When crossing *C. tshokwe* (a medium duration EOD species, 5 ms) with two short EOD duration species (*C. compressirostris* and *C. tamandua*), the hybrids showed EOD elongation during ontogeny. This is probably due to the upregulation of genes in the adult electric organ of *C. tshokwe*, in particular voltage-gated potassium channel genes which have been hypothesized to be up-regulated in the hybrids, even during the early ontogenetic stages (Kirschbaum et al. 2016; Nagel et al. 2017).

Hybrids of *C. rhynchophorus* and other species showed an intermediate duration EOD. However, the EOD is occasionally closer to *C. rhynchophorus*, e.g. in the hybrid *C. compressirostris* x *C. rhynchophorus* (Kirschbaum et al. 2016). The pronounced surface proliferations (papillae), which characterized the two species with extremely elongated EOD (*C. rhynchophorus* and *C. numenius*), is also observed in the hybrids when *C. rhynchophorus* is one of the parents (Kirschbaum et al. 2016). However, the adult EOD in those hybrids does not reach the same long duration as in adult *C. rhynchophorus*. In hybrid *C. compressirostris* x *C. rhynchophorus*, the adult EOD is very close to a juvenile EOD in *C. rhynchophorus* in both shape and waveform (Kirschbaum et al. 2016; Nguyen et al. 2020).



♀ Parent	♂ Parent	Hybrid
75 Cc  A1	80/87 Ct  A2	96  A3 165 mm
80 Cts  B1	80/87 Ct  B2	101/104  B3 149 mm
52 Cr (P)  C1	80/87 Ct  C2	67/71 P  C3 98 mm
80 Cts  D1	75 Cc  D2	78/89  D3 135 mm
52 Cr (P)  E1	75 Cc  E2	59/63 P  E3 85 mm
52 Cr (P)  F1	80 Cts  F2	65 P  F3 107 mm
52 Cr (P)  G1	60 Cn (P)  G2	60 P  G3 92 mm

**Figure 1.4** Intragenus hybridization in *Campylomormyrus* species. The electric organ structure of both parents and the hybrid are shown here (S, stalk; E, electrocyte; P, papillae). CC, *C. compressirostris*; Ct, *C. tamandua*; Cts, *C. tshokwe*; Cr, *C. rhynchophorus*; Cn, *C. numenius*. This figure is modified from **Kirschbaum et. al 2016**.

### 1.6 Aims of this study

The diversification of the EOD is a potentially important mechanism underlying adaptive radiations in *Campylomormyrus*. So far, several studies have investigated potential mechanisms of

the EOD variations at histological, electrophysiological and genetic levels in this genus (Lamanna et al. 2015; Paul et al. 2015; Nagel et al. 2017; Korniienko et al. 2021). However, the basis behind the EOD variation among species is still only partially understood, especially EOD duration diversification.

Intragenus hybrids have special EOD traits (Kirschbaum et al. 2016). Interestingly, whenever we crossed parental species between short EOD and elongated EOD, the hybrids express elongated EOD compared to the short EOD parent species. I am interested to understand why the EOD bias occurred in the intragenus hybridization. Do alleles from elongated EOD parental species have higher expression than short EOD parental species?

The development of EOD in *Campylomormyrus* is also not well understood. The EOD of short duration species *C. compressirostris* exhibits no change during ontogeny, while species with elongated adult EOD show multiple different juvenile EODs until they reach their adult EOD. What causes the consistency and change during EOD development?

The main focus of this study is on the exploration of the genetic basis of the diversification in EOD of different species/hybrids and life stages (ontogeny) in the African weakly electric fish genus *Campylomormyrus*, as well as the EOD bias in intragenus hybrids. Therefore, I started with the generation of a high quality genome in this genus. My study consists of the three following parts:

1. **Genome assembly on species *C. compressirostris*.** I generated a high quality whole genome for *C. compressirostris* using Pacbio long read sequencing and evidence-based genome annotation. Based on this, a gene family analysis was performed to explore gene family evolution in different teleost fish lineages after whole genome duplication.

2. **Transcriptomic comparisons among different adult species/hybrids.** I used mRNA sequencing technology to explore the gene expression pattern between electric organ and skeletal muscle in different adult *Campylomormyrus* species/hybrids. The gene expression pattern of species with different EOD duration was also utilized to identify genes potentially involved in the regulation of EOD waveform. In addition, allele specific expression analysis was performed to investigate the biallelic expression imbalance in the hybrids.
3. **Gene expression pattern during ontogeny.** For the purpose of identifying genes involved in the EOD development during ontogeny, I compared the gene expression of electric organs between juveniles and adults in two species and their hybrid. To further understand the phenotypic evolution during ontogeny in the hybrid, the allele specific expression was compared to explore the biallelic expression difference between juvenile and adult hybrids.

## 2 Summary of articles

### 2.1 Summary of Article I

**A new genome assembly of an African weakly electric fish (*Campylomormyrus compressirostris*, Mormyridae) indicates rapid gene family evolution in Osteoglossomorpha.**

Cheng, F., Dennis, A. B., Osuoha, J. I., Canitz, J., Kirschbaum, F., & Tiedemann, R. (2023). *BMC Genomics*, 24(1), 129.

A well-annotated genome in *Campylomormyrus* genome is imperative to understand the convergent evolution in electric fish and the adaptive radiation in mormyrids. An 862 Mb size genome was generated for the species *C. compressirostris* using Pacbio long read sequences. There were 34,492 protein-coding genes predicted in 1,479 contigs, which is a noteworthy higher number than in the two other available annotated Osteoglossomorpha genomes (*Paramormyrops kingsleyae* and *Scleropages formosus*). A gene family evolution analysis via the program CAFE5 was employed to analyze the gene family expansion and contraction in representative genomes from Osteoglossomorpha, Otomorpha and Euteleosteomorpha. Based on 33 teleost fish genomes, the Osteoglossomorpha showed an overall faster gene family turnover rate compared to Otomorpha and Euteleosteomorpha. In addition, Osteoglossomorpha exhibited significantly higher ratios of expanded/contracted gene family numbers compared to the other two groups. We manually curated 16 genes from the *Kv1* subfamily, with two tandem duplicated gene copies of *KCNA7a*. The significantly higher ratios of expanded/contracted gene family numbers and the high number of *Kv1* subfamily genes suggested that the basal taxon Osteoglossomorpha might have a faster gene family evolutionary speed. This knowledge will help to improve our understanding the evolution of electric fish and Osteoglossomorpha taxa.

**Authors contribution:** I performed all the lab work, the analyses and manuscript writing with the input from R. Tiedemann, A. B. Dennis, J. Canitz, F. Kirschbaum. R. Tiedemann conceived and supervised the study. J. I. Osuoha curated the *Kv1* genes with the input from me and J. Canitz. F. Kirschbaum took part of the supervision.

## 2.2 Summary of Article II

### **Gene and allele specific expression underlying the electric signal divergence in African weakly electric fish.**

Cheng, F., Dennis, A. B., Baumann, O., Kirschbaum, F., Abdelilah-Seyfried, S., & Tiedemann, R. In submission at *Communications Biology*.

Electric fishes have independently evolved at least six times. The weakly electric fish clade of mormyrids is one of the most diverse family of freshwater fishes. They possess a specific myogenic electric organ (EO) derived from skeletal muscle (SM) fibers. The magnificent adaptive radiation possibly resulted from their striking divergence in electric organ discharge (EOD) among species, which is considered to contribute to pre-zygotic isolation. I sequenced the mRNA of EOs and SMs from the species *Campylomormyrus compressirostris* (0.4 ms duration EOD), *C. tshokwe* (5 ms duration EOD), *C. rhynchophorus* (40 ms duration EOD) and two cross-species hybrids *C. compressirostris* ♂ x *C. tshokwe* ♀ (0.5 ms duration EOD), and *C. compressirostris* ♂ x *C. rhynchophorus* ♀ (4 ms duration EOD).

There were 1,444 genes up-regulated genes in the EO compared to SM that were shared by all five species/hybrids cohorts. Several genes, e.g. *SCN4aa*, *KCNA7a*, *SIX2a*, *HEY1* and several isoforms of sodium/potassium ATPase  $\alpha$  and  $\beta$  subunits, showed convergent expression pattern in the EO of different studied electric fish lineages. The differentially expressed actin-related genes (F-actin-

related genes, unconventional myosin genes and *MEF2b*) suggested that the developmental transition in the EO might be different across mormyrids. In addition, we also identified a potential transcription factor, *KLF5*, which might drives the expression of regulating potassium channels in the EO of *Campylomormyrus*.

We made cross species comparisons among purebred species and two different tissues to investigate the gene expression relative to EOD duration variation. Three types of EOD-duration-related gene expression pattern were identified. The up-regulation of genes *KCNJ2* and *KLF5* as well as the down-regulation of genes *KCNK6* and *KCNQ5* might contribute to the increased EOD duration. The allelic imbalanced expression at the *KCNJ2* gene in hybrid *C. compressirostris* x *C. rhynchophorus*, i.e., a higher expression of the *C. rhynchophorus* allele, points towards a cis-regulatory difference at the locus. It supports our hypothesis that the gene *KCNJ2* might be a powerful candidate for the EOD duration modulation.

**Authors contribution:** I performed RNA isolation, cDNA library preparation, gene expression and allele specific analyses, and manuscript writing with input from R. Tiedemann, A.B. Dennis, O. Baumann, F. Kirschbaum and S. Abdelilah-Seyfried. R. Tiedemann conceived and supervised this study, and provided financial and logistical support. F. Kirschbaum partially supervised the project and provided biological information about electric fish. A.B. Dennis contributed to the transcriptome analyses.

### **2.3 Summary of Article III**

**Gene and allele specific expression during electric organ ontogeny in African weakly electric fish and their hybrids.**

Cheng, F., Dennis, A. B., Baumann, O., Kirschbaum, F., Domínguez, M., & Tiedemann, R. 2023.  
To be submitted.

Hybridization can contribute to our understanding of the evolution of complex phenotypes. The African weakly electric fish genus *Campylomormyrus* possess a myogenic electric organ (EO) to produce an electric organ discharge (EOD) which remarkably varies across species/hybrids and throughout ontogeny. The EOD of *C. compressirostris* does not change during ontogeny, while the EOD of *C. rhynchophorus* starts from a similar early juvenile EOD and reaches an elongated EOD in multiple intermediate stages. The EOD development in the hybrid across *C. compressirostris* and *C. rhynchophorus* shows early juvenile EOD identical to *C. compressirostris* and eventually reaches an EOD of intermediate length (around 4 ms) between the parental species after multiple ontogenetic stages. Interestingly, the adult EOD in the hybrid resembles a juvenile EOD of *C. rhynchophorus*, indicating the EOD development in this hybrid is closer to *C. rhynchophorus* than to *C. compressirostris*.

We performed pairwise comparison of gene expression analysis between juvenile and adult EO of *C. compressirostris* (EOD duration 0.4 ms in juveniles and 0.4 ms in adults), *C. rhynchophorus* (EOD duration 5 ms in juveniles and 40 ms in adults) and their hybrids *C. compressirostris* ♂ x *C. rhynchophorus* ♀ (EOD duration 0.4 ms in juveniles and 4 ms in adults). Differentially expressed genes between juvenile and adult EOs were significantly enriched in “membrane”, “plasma membrane” and “cytoplasm” Go Ontology terms. Candidate genes potentially contributing to the EOD development of *C. rhynchophorus* and the hybrids were identified, e.g. *ADCYAPI*, *KCNJ2*. Other up- or down-regulated potassium channel genes might also contribute to regulate the EOD development.

In addition, allele specific expression analysis was performed to investigate the allelic expression during ontogeny in hybrids. The allele specific expression showed a general dominance of the *C. compressirostris* allele in the hybrids at both juvenile and adult life stages. Only the gene *KCNJ2* exhibited a dominant expression of the *C. rhynchophorus* allele and was increasingly dominant from juvenile to adult stages. This suggests that the EOD development in hybrids could be related to the increasing allelic expression of the *C. rhynchophorus* allele of this gene under a scenario of an overall dominance of *C. compressirostris* alleles.

**Author contributions:** R. Tiedemann conceived and supervised this study, and provided financial and logistical support. I performed RNA isolation, cDNA library preparation, gene expression and allele specific analyses, and manuscript writing with input from R. Tiedemann, A.B. Dennis, O. Baumann, F. Kirschbaum and M. Domínguez. F. Kirschbaum partially supervised the project and provided biological information about electric fish.



### 3 Article I (published)

**A new genome assembly of an African weakly electric fish (*Campylomormyrus compressirostris*, Mormyridae) indicates rapid gene family evolution in Osteoglossomorpha.**

Feng Cheng<sup>1</sup>, Alice B. Dennis<sup>1,2</sup>, Josephine Ijeoma Osuoha<sup>1</sup>, Julia Canitz<sup>3</sup>, Frank Kirschbaum<sup>1,4</sup>,  
Ralph Tiedemann<sup>1,\*</sup>

1 Unit of Evolutionary Biology and Systematic Zoology, Institute of Biochemistry and Biology,  
University of Potsdam, Potsdam, Germany

2 Laboratory of Adaptive Evolution and Genomics, Research Unit of Environmental and  
Evolutionary Biology, Institute of Life, Earth & Environment, University of Namur, Namur,  
Belgium

3 Senckenberg German Entomological Institute, Müncheberg, Germany

4 Department of Crop and Animal Science, Faculty of Life Sciences, Humboldt University, Berlin,  
Germany

#### **Abstract**

**Background:** Teleost fishes comprise more than half of the vertebrate species. Within teleosts, most phylogenies consider the split between Osteoglossomorpha and Euteleostei/Otomorpha as basal, preceded only by the derivation of the most primitive group of teleosts, the Elopomorpha. While Osteoglossomorpha are generally species poor, the taxon contains the African weakly electric fish (Mormyroidei), which have radiated into numerous species. Within the mormyrids, the genus *Campylomormyrus* is mostly endemic to the Congo Basin. *Campylomormyrus* serves as a model to understand mechanisms of adaptive radiation and

ecological speciation, especially with regard to its highly diverse species-specific electric organ discharges (EOD). Currently, there are few well-annotated genomes available for electric fish in general and mormyrids in particular. Our study aims at producing a high-quality genome assembly and to use this to examine genome evolution in relation to other teleosts. This will facilitate further understanding of the evolution of the osteoglossomorpha fish in general and of electric fish in particular.

**Results:** A high-quality weakly electric fish (*C. compressirostris*) genome was produced from a single individual with a genome size of 862Mb, consisting of 1,497 contigs with an N50 of 1,399 kb and a GC-content of 43.69%. Gene predictions identified 34,492 protein-coding genes, which is a higher number than in the two other available Osteoglossomorpha genomes of *Paramormyrops kingsleyae* and *Scleropages formosus*. A Computational Analysis of gene Family Evolution (CAFE5) comparing 33 teleost fish genomes suggests an overall faster gene family turnover rate in Osteoglossomorpha than in Otomorpha and Euteleostei. Moreover, the ratios of expanded/contracted gene family numbers in Osteoglossomorpha are significantly higher than in the other two taxa, except for species that had undergone an additional genome duplication (*Cyprinus carpio* and *Oncorhynchus mykiss*). As potassium channel proteins are hypothesized to play a key role in EOD diversity among species, we put a special focus on them, and manually curated 16 *Kv1* genes. We identified a tandem duplication in the *KCNA7a* gene in the genome of *C. compressirostris*.

**Conclusions:** We present the fourth genome of an electric fish and the third well-annotated genome for Osteoglossomorpha, enabling us to compare gene family evolution among major teleost lineages. Osteoglossomorpha appear to exhibit rapid gene family evolution, with more gene family expansions than contractions. The curated *Kv1* gene family showed seven gene clusters,

which is more than in other analyzed fish genomes outside Osteoglossomorpha. The *KCNA7a*, encoding for a potassium channel central for EOD production and modulation, is tandemly duplicated which may be related to the diverse EOD observed among *Campylomormyrus* species.

**Keywords:** *Campylomormyrus*; Pacbio sequencing; gene family; Osteoglossomorpha, *Kv1*

## **Background**

Teleost fishes comprise more than half of the vertebrate species in the world, showing a marvelous biodiversity concerning morphology, ecology and behavior [1]. It has been shown that a teleost-specific whole genome duplication (TS-WGD) had occurred in the common ancestor of all extant teleost [2–5]. Although there is no solid evidence to support the connection between the TS-WGD and the successful radiation of teleosts, the former provided enormous opportunities for gene innovation and evolution [6]. The redundant duplicated genes may be free to evolve new or related functions in the course of long-term modification and divergence, and may hence have fostered functional and phenotypic diversification in teleost fish [7].

One of the possible trajectories of diversification following gene duplication is parallel evolution among disparate taxa [8], as exemplified in the evolution of electric organs in unrelated lineages [9]. Among fish, myogenic electric organs have independently evolved at least six times, enabling the generation of electric fields, which are used for communication, navigation, and in extreme cases for predation and defense [10–14]. This electric organ-specific parallel evolution appeared both in elasmobranch fish and two unrelated teleost lineages: the Gymnotiformes from South America and the Mormyroidei from Africa [10].

The vast majority of African weakly electric fishes belongs to the Mormyridae, one of the most diverse family of freshwater fishes. They are endemic to Africa where there are at least 188 described species [15]. The genus *Campylomormyrus* comprises 15 described species, most endemic to the Congo Basin [15, 16]. As in other mormyrids, the electric organ of *Campylomormyrus* is derived from myogenic tissue and located in the caudal peduncle [17]. It is composed of specialized electrocytes which produce externally measurable electric organ discharges (EODs) [17]. The species-specific EOD displays a huge diversity in signal duration and waveform [18, 19]. However, the proximate mechanisms underlying the species divergence of EOD among species are only partially understood [20–23]. In order to better understand the evolution of this genus, a high-quality genome is imperative. Up to now, three complete genomes have been published of electric fishes of the genera *Paramormyrops*, *Electrophorus* and *Brachyhypopomus* [24–26]. Hence, our genomic knowledge is still too incomplete for a comprehensive assessment of electric fish's molecular evolution and its impact on phenotypic divergence.



**Fig. 1:** Photo and typical electric organ discharge (EOD, upper left corner) of the sequenced species *Campylomormyrus compressirostris*. (Photo taken by Frank Kirschbaum)

The aim of our study is to generate a high-quality genome for the African weakly electric fish species *Campylomormyrus compressirostris*, a species that produces a biphasic pulse type EOD

(Fig. 1). This genome will provide an invaluable resource for the genus *Campylomormyrus*, an established model for adaptive radiation and ecological speciation [27]. As a first step, we here use this genome to compare the evolution of gene family size in *C. compressirostris* relative to other electric fishes, and to teleost fish in general. In addition, we have manually curated and examined the important *Kv1* voltage-gated potassium channel genes, which is hypothesized to be involved in the diversification of the EOD signal and speciation in weakly electric fish [19].

## Results

### Genome assembly of *Campylomormyrus compressirostris*

Here we report a new genome assembly from the African weakly electric fish *C. compressirostris*. The specimen used for sequencing was artificially bred and raised at the University Potsdam, Germany. A total of 294 Gb Pacbio raw data (~294.3 billion reads) was generated. Circular consensus sequencing (CCS) produced 15.5 Gb (~1.03 million reads) high fidelity (HiFi) raw data. The produced HiFi data were analyzed based on their k-mer distribution [28] to estimate the genome size (799 Mb) and genome heterozygosity (0.96%).

Using the hifiasm assembler [29], the final assembly is 862Mb in size, and contains 1,497 contigs with a contig N50 of 1.3Mb and a GC-content of 43.69% (Table 1). The largest contig has a length of 1,399kb. The assembly also produced the set of alternate contigs (815.8Mb). The genome quality was close to that of the *Brachyhyppopomus occidentalis* genome, and significantly improved compared with the two other published electric fish genomes: *Electrophorus electricus* and *Paramormyrops kingsleyae* (Table 1).

**Table 1** Comparison of available genome assemblies for 4 electric fish (Osteoglossomorpha: *C. compressirostris*, *P. kingsleyae*; Gymnotiformes: *E. electricus*, *B. occidentalis*) and 1 non-electric osteoglossomorph fish (*S. formosus*).

	Osteoglossomorpha			Gymnotiformes	
	<i>C. compressirostris</i>	<i>P. kingsleyae</i>	<i>S. formosus</i>	<i>E. electricus</i>	<i>B. occidentalis</i>
Sequencing Platform	Pacbio HiFi	Illumina HiSeq2000	Illumina HiSeq2000	Illumina HiSeq2000	10 x
Genome size (Mb)	862	880	779	720	540.3
Coverage	14x	83x	137.6x	55x	46x
Complete BUSCOs	94.6%	95.0%	-	97.0%	93.8%
n contigs	1,497	4,496	-	340,589	1,435
Contig N50 (kb)	1,399	37.6	30.73	104	5,400
GC content	43.9%	43.0%	-	42.5%	44.6%
Genes Predicted	34,492	27,677	22,016	22,000	34,347

The integrity of the assembly was demonstrated by 94.6% Benchmarking Universal Single-Copy Orthologs (BUSCO) [30] completeness, indicating the high degree of completeness of the gene regions.

### Genome annotation

Genome annotation was conducted in several steps. First, repeats were identified and masked. The repeat content was identified based on the RepeatModeler [31], and accounted in total for 27.28% (235.37Mb) of the assembled genome. Next, gene predictions were made using combined evidence from empirical transcriptomic data and protein references from the National Center for Biotechnology Information (NCBI). These were provided to the MAKER pipeline, which predicted 34,492 protein-coding genes, 280,886 exons and 246,394 introns (Table 2). The coding sequence (CDS) covers 5.3% of the genome. Over 90% of the genes have an annotation edit

distance (AED) of 0.5 or lower (Additional file 1), suggesting that they are well supported by either protein or RNA-seq evidence. The number of predicted protein-coding genes is notably higher than in the other two sequenced Osteoglossiformes fishes: *P. kingsleyae* and *Scleropages formosus* which had 27,677 and 22,016 protein-coding genes, respectively (Table 1).

**Table 2** Genome annotation statistics

	Exons	Introns	Genes	CDS
Number	280,886	246,394	34,492	34,492
Longest in kb	26.4	292.8	424	534
Mean length	225	1,095	9,645	1,330
% genome covered by			38.6	5.3

### Orthogroup identification in teleost fish

The gene family analyzer CAFE5 (Computational Analysis of gene Family Evolution) [32] was used to compare annotated gene content in our genome with 33 genomes of teleost fish that we selected based on contiguity and taxonomic representation (Additional file 2).

**Table 3** Number of orthogroups (OGs) with significant changes in gene number ( $p < 0.05$ ) among teleosts, compared to all OGs in teleost fish. Overrepresentation of certain functions was tested with Fisher's exact test.

	OGs with significantly changed gene number	All OGs	P-value
Total number	368	23,613	
Number with annotation	276	20,663	
Number associated with zinc finger proteins	12	809	0.6416
Number associated with transposons	10	36	<0.001
Number associated with immunoglobulins	10	90	<0.001
Number associated with GTPases	7	202	0.0228

Orthogroups (OGs) were clustered among the filtered peptides sequences in OrthoFinder [33]. We obtained 23,613 OGs from teleost fish by OrthoFinder, of which 402 were identified as single copy. There are 500 unique OGs in *C. compressirostris*, and 919 in Mormyridae (represented by *C. compressirostris* and *P. kingsleyae*). A total of 1,169 unique OGs were identified among Osteoglossomorpha, 1,134 among Otomorpha, and 2,540 among Euteleosteomorpha.

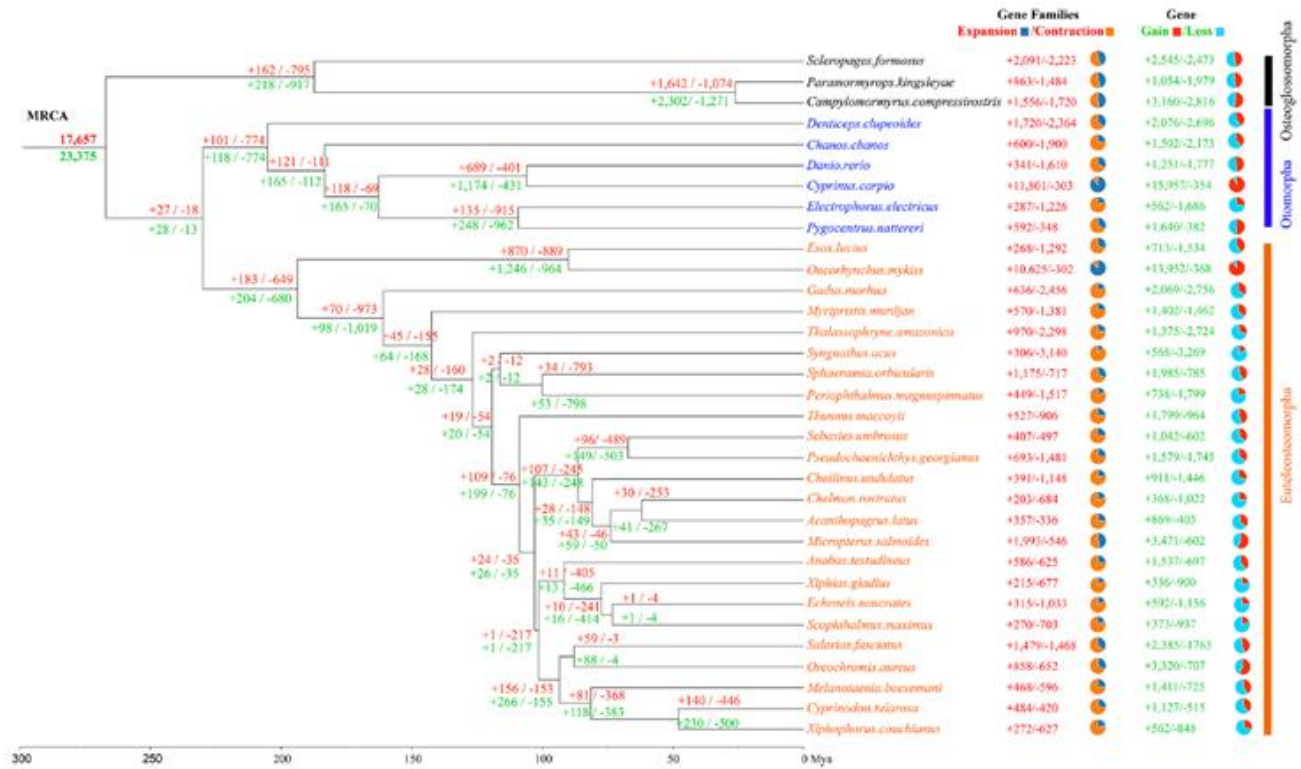
### **Gene and gene family expansion and contraction analysis**

We estimated both expansions and contractions in gene family size across the evolutionary history of all teleosts. All the 23,613 OGs from OrthoFinder were used as input in CAFE5. CAFE5 estimated the gene family turnover rate lambda for each group of Osteoglossiforpha, Otomorpha and Euteleosteomorpha.

Based on the gene family clustering results in CAFE5, 368 OGs were significantly changed in gene numbers per family among teleost, of which 276 were annotated using the UniProt database (Table 3). From this set, OGs were repeatedly (over 5 times) associated with zinc finger protein, transposon, immunoglobulin and GTPase. We put the relative frequency of OGs with these functions into perspective of their occurrence among the 20,663 annotated OGs in total, employing Fisher's exact tests [34]. Among the OGs with significantly changed gene number contents across teleost lineages, OGs related to transposons, immunoglobulins and GTPases are significantly overrepresented. In the CAFE5 gene family analysis, the estimated gene family turnover rate lambda was larger for Osteoglossomorpha (0.0029) than for Otomorpha (0.0022) and Euteleosteomorpha (0.0022).

In order to eliminate the bias introduced by species that have undergone an additional and more recent WGD, we repeated the CAFE5 analysis without those species (i.e., excluding *Cyprinus*

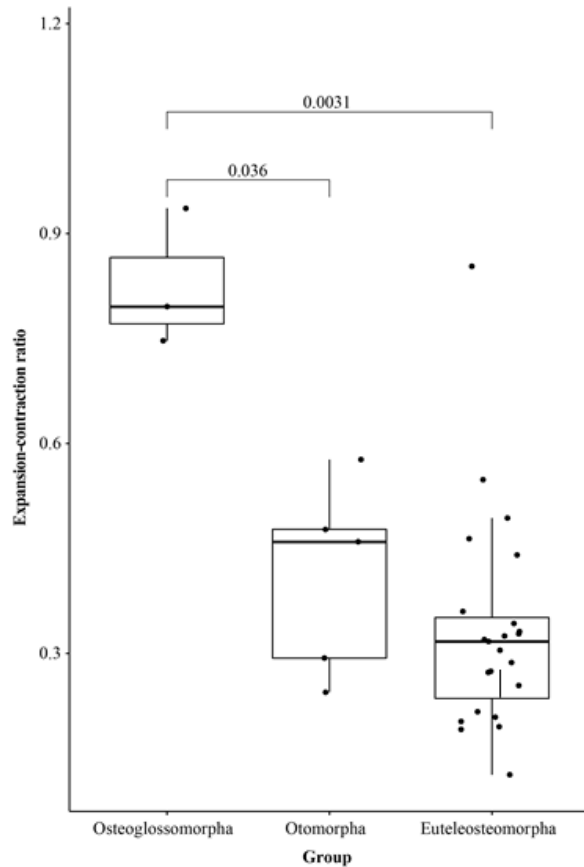




**Fig. 2:** Inferred MRCA (most recent common ancestor) numbers of genes (green) and gene families (red) as well as the expansions (gains, +) and contractions (losses, -) in genomes of different teleost lineages. The pie charts show the gene/gene family expansions and contractions of species compared to the MRCA.

*carpio* and *Oncorhynchus mykiss*). With this removal, Osteoglossomorpha (0.0030) show an even larger lambda, relative to Otomorpha (0.0020) and Euteleostomorpha (0.0019), indicating a higher gene family turnover (birth-death) rate in Osteoglossomorpha.

We compared the inferred gene and gene family change for each node, relative to its preceding node in the phylogeny. 1,556 gene families were expanded in *C. compressirostris* whereas 1,720 contracted (Fig. 2). This species gained 3,160 genes and lost 2,816 genes. The common ancestor of *C. compressirostris* and *P. kingsleyae* had 1,642 gene families expanded and 1,074 contracted, *P. kingsleyae* only had inferred expansion in 863 gene families, relative to contraction 1,484 gene families (Fig. 2).



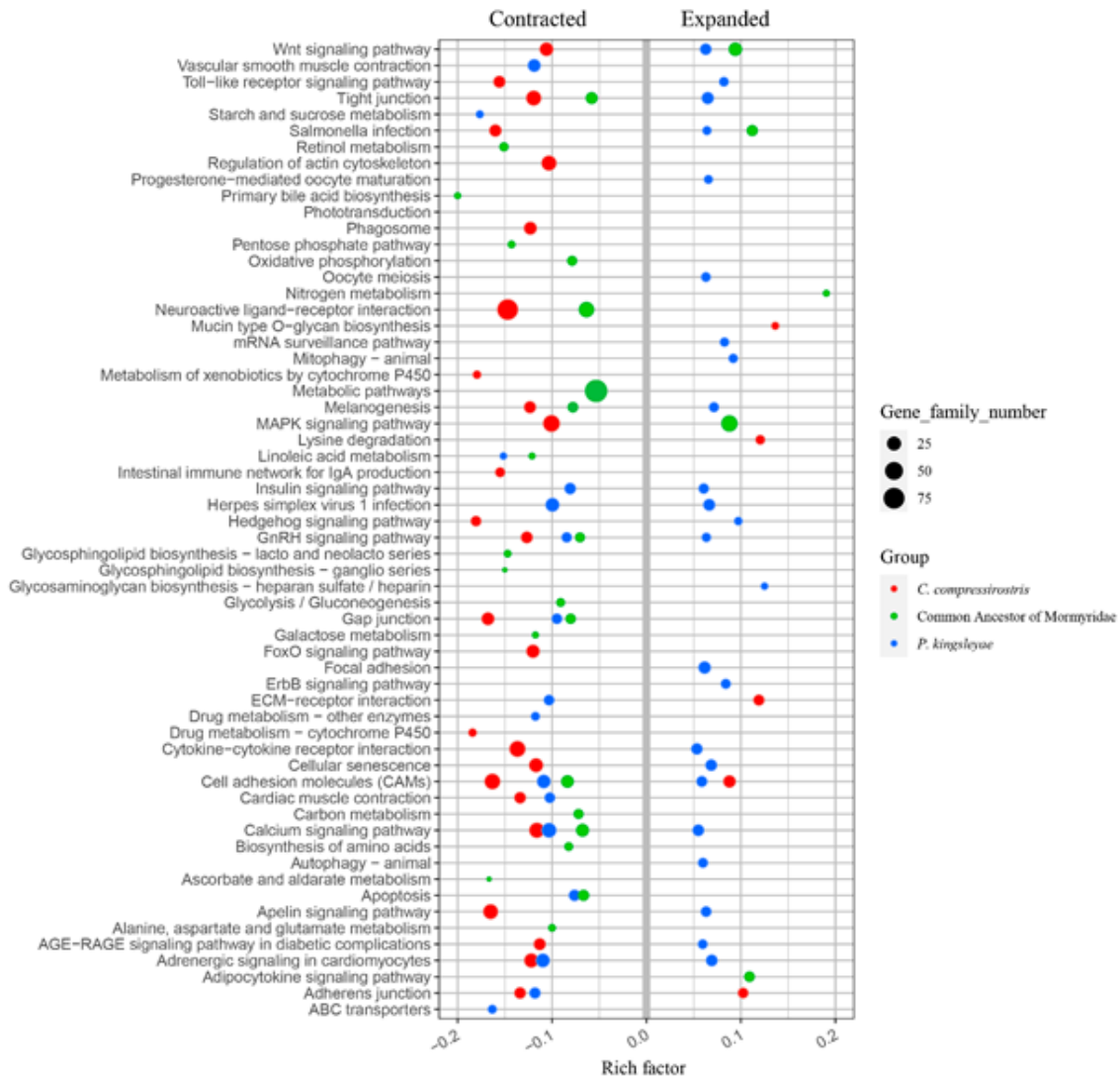
**Fig. 3:** Box-and-scatter plot of gene family expansion/contraction ratios in the groups Osteoglossomorpha, Otomorpha and Euteleostomorpha. The values above box plot are the P-values between corresponding groups from a t-test.

For comparing the lineage-specific gene family and gene change relative to their most recent common ancestor (MRCA) of all selected teleost fishes, we summarized the gene and gene family change for every species (Fig. 2, Additional file 3) and counted the ratios of expanded/contracted gene family (as well as gained/lost gene) numbers. In most of the analyzed species, more gene families contracted than expanded, with the exception of the two species with additional, recent genome duplications: *C. carpio* and *O. mykiss*. Leaving these two species out, the ratio of gene family expansions/contractions was significantly higher in Osteoglossomorpha than in Otomorpha and Euteleostomorpha (p-values of 0.036 and 0.0031 respectively, t-Test [35], Fig. 3).

### Gene families and pathways with increased turnover in electric fish and Osteoglossomorpha

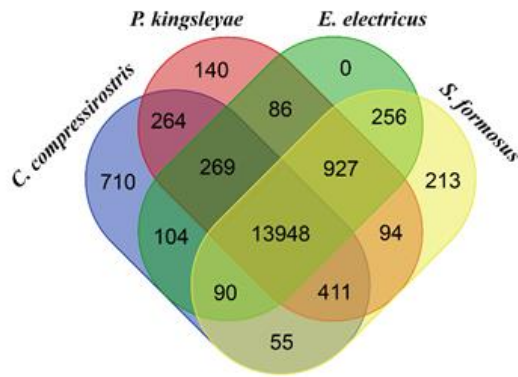
We assigned gene families to Kyoto Encyclopedia of Genes and Genomes (KEGG) [36] pathways to identify those pathways exhibiting a significantly ( $p < 0.05$ ) elevated turnover (i.e., a significantly higher number of either contracted and expanded gene families) in the two mormyrids *C. compressirostris* and *P. kingsleyae*, relative to their common ancestor. This analysis yielded 60 significantly enriched pathways (Fig. 4), of which 25 contained contracted OGs and 5 expanded OGs in *C. compressirostris*; for *P. kingsleyae*, 16 pathways comprised contracted OGs and 24

expanded ones. For the ancestor node of both mormyrid fishes, 22 pathways with elevated turnover exhibit contracted OGs and 5 expanded ones. The rich factor indicates the degree of the enrichment in the respective KEGG pathways (Fig. 4). The pathway with highest rich factor is primary bile acid biosynthesis in contracted OGs, and is nitrogen metabolism in expanded OGs.



**Fig. 4:** Numbers of expanded and contracted gene families in KEGG pathways with significantly elevated turnover (contraction or expansion) among *C. compressirostris* (red), *P. kingsleyae* (blue) and their common ancestor (green). The plot size represents the gene family number in the respective species. Note that non-significant values are not plotted, hence not all pathways have dots for all taxonomic groups.

To examine shared OGs, a VennDiagram [37] was created to visualize all OGs in three electric fish (*C. compressirostris*, *P. kingsleyae* and *E. electricus*) and one non-electric fish *S. formosus* from Osteoglossomorpha (Fig. 5). There were 269 enriched OGs shared among the electric fishes,



**Fig. 5:** Venn Diagram graph of all orthologous gene families shared/not shared among four species (*C. compressirostris*, *P. kingsleyae*, *E. electricus*, *S. formosus*).

and 411 enriched OGs shared among Osteoglossomorpha. 264 enriched OGs were only shared among mormyrids. Although this is a small dataset, it could suggest that OGs turnover patterns are more similar among phylogenetically related groups (here, osteoglossomorphs) than among species having convergently evolved an active electric sense.

### **KCNA gene cluster curation**

The potassium voltage-gated channel subfamily A (*KCNA*, *Kv1*) encodes shaker-related voltage-gated potassium channels, which are considered as a component of electric organ discharges. 16 complete *Kv1* genes, which contained both start and stop codons, were manually curated in the *C. compressirostris* genome (Table 4, Additional file 4). 11 of them were predicted in the annotation pipeline. Manual searches identified *KCNA3a/b*, *KCNA7b* and *KCNA10a/b*. We could not find *KCNA5a* gene, which was considered to be lost according to the available resources.

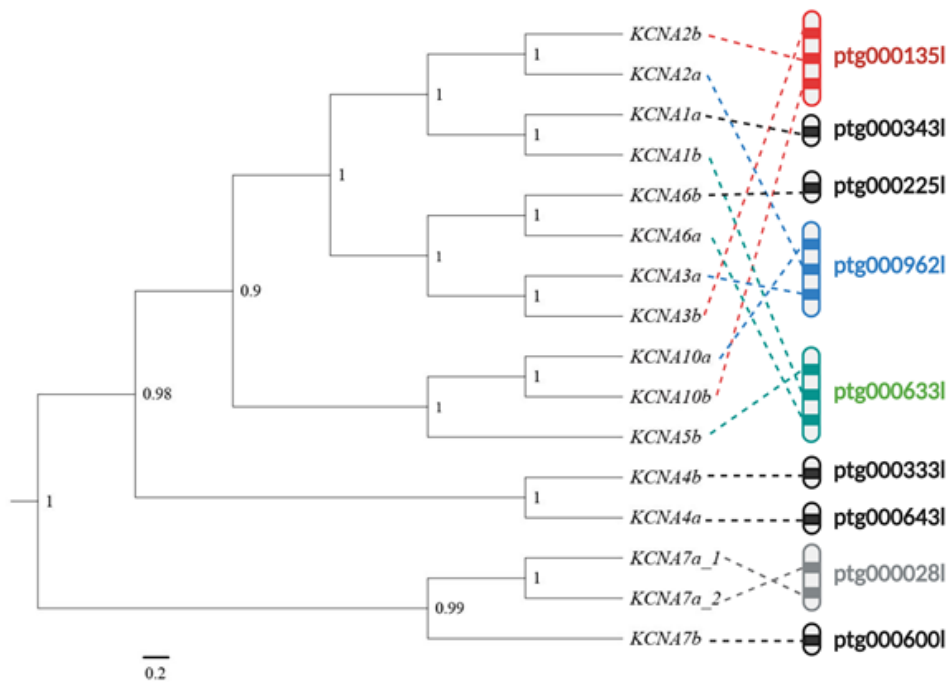
The CDS length among *KCNA* genes varied from ~1,400 bp to ~2,000 bp. We detected two *KCNA7a* gene copies in contig ptg000028l with a regional distance of ~26kb (Table 4). Genes *KCNA1a/b*, *KCNA2a/b*, *KCNA3a/b*, *KCNA4a/b*, *KCNA5b*, *KCNA6a/b* and *KCNA10a/b* have only

**Table 4** *KCNA* genes information in the *C. compressirostris* genome

Gene	Contig	Start position	End position	Number of exons	Length of cds
<i>KCNA1a</i>	ptg0003431	158,599	160,074	1	1,476
<i>KCNA1b</i>	ptg0006331	255,2072	2,553,541	1	1,470
<i>KCNA2a</i>	ptg0009621	2,034,837	2,036,324	1	1,488
<i>KCNA2b</i>	ptg0001351	6,635,624	6,637,102	1	1,479
<i>KCNA3a</i>	ptg0009621	2,057,829	2,059,386	1	1,557
<i>KCNA3b</i>	Ptg0001351	6,613,381	6,614,937	1	1,557
<i>KCNA4a</i>	ptg0006431	775,778	777,787	1	2,010
<i>KCNA4b</i>	ptg0003331	510,057	512,045	1	1,989
<i>KCNA5b</i>	ptg0006331	2,507,887	2,509,569	1	1,683
<i>KCNA6a</i>	ptg0006331	2,583,269	2,584,375	1	1,107
<i>KCNA6b</i>	ptg0002251	702,207	703,643	1	1,437
<i>KCNA7a_1</i>	ptg0000281	8,120,428	8,129,001	2	1,539
<i>KCNA7a_2</i>	ptg0000281	8,089,350	8,094,209	2	1,542
<i>KCNA7b</i>	ptg0006001	1,376,337	1,378,893	2	1,551
<i>KCNA10a</i>	ptg0009621	2,004,772	2,006,448	1	1,677
<i>KCNA10b</i>	ptg0001351	6,670,889	6,672,565	1	1,677

one exon, whereas *KCNA7a\_1*, *KCNA7a\_2* and *KCNA7b* were found to have two exons. Among the newly discovered duplications of *KCNA7a*, the exon1 of *KCNA7a\_1* and *KCNA7a\_2* are identical, however, there are 55 single nucleotide polymorphisms (SNPs) among them in the 855 bp of exon2. The p-distance was 0.0657 between two *KCNA7a* copies of exon2.

The phylogenetic analysis of all *Kv1* genes in BEAST v1.8.4 [38] suggests a basal position of *KCNA7a/b* genes (Fig. 6). The *KCNA1a/b* and *KCNA2a/b* form a monophyletic cluster, as do *KCNA5b* and *KCNA10a/b*, *KCNA3a/b* and *KCNA6a/b*. Those monophyletic gene clusters corroborate the hypothesis that the three clusters resulted from a complete duplication of the original cluster instead of independent tandem duplication [39].



**Fig. 6:** Bayesian tree of all curated *Kvl* genes in *C. compressirostris* genome. The posterior probability value is shown on each node. Genes are also mapped to their respective contig. The following genes are linked, i.e., mapped to the same contig: I: *KCNA2b*, *KCNA3b*, *KCNA10b*; II: *KCNA2a*, *KCNA3a*, *KCNA10a*; III: *KCNA1b*, *KCNA5b*, *KCNA6a*; IV: *KCNA7a\_1*, *KCNA7a\_2*.

## Discussion

### Genomic resources for electric fish and early teleost evolution

This is the fourth genome of an electric fish, and the third well-annotated genome for the basal teleost taxon Osteoglossomorpha. Its quality is significantly improved with regard to contig length, when compared to the existing genomes for electric fish (*P. kingsleyae*, *E. electricus* [25, 26] and the non-electric Osteoglossomorpha fish *S. formosus* [40]). Regarding the number of contigs, completeness of BUSCO matches, and GC-content, our genome is also comparable to the recently released *B. occidentalis* genome ([24]; a gymnotiform electric fish), which was generated by 10x genomics linked read sequencing (Table 1). Our new genome will provide a valuable resource for

future research on the evolution and ecology not only of electric fish, but also of the basal Osteoglossomorpha and early teleosts in general.

The draft annotation from MAKER predicted 34,492 protein-coding genes (Table 1). This is a substantially higher number than revealed in the annotations of *P. kingsleyae* (27,677) and *S. formosus* (22,016) [25] in the original publications. It had been hypothesized that mormyrid fishes may have a larger number of genes than other non-electric osteoglossiformes [25], however, gene counts could also vary due to annotation artifacts, for example if reads from single genes are erroneously annotated to two genes [25], or uncollapsed haplotypes in the assembly. The annotations we used in our CAFE5 analysis were downloaded from NCBI pipeline, which resulted in 23,862 and 23,537 genes for *P. kingsleyae* and *S. formosus*, respectively, hence not supporting a generally increased gene number in mormyrids, relative to other Osteoglossomorpha fishes. To investigate whether the larger number of genes in *C. compressirostris* is due to fragmentation of single genes in the annotation from MAKER, we mapped the CDS of *C. compressirostris* to the reference *P. kingsleyae* CDS from NCBI using MUMmer4.0 [41], allowing the *C. compressirostris* only map to the best hit in the reference (Additional file 5). This showed 3,076 overmapped genes in *C. compressirostris* that could indicate more than one gene model matching to a single gene model in *P. kingsleyae*. However, these could also be matches between similar genes and true duplications. In addition, there were an additional 10,846 and 2,939 unique genes in *C. compressirostris* and *P. kingsleyae*, respectively. These did not map between the species and suggests an excess of new and expanded genes in *C. compressirostris*, or lost/un-annotated genes in *P. kingsleyae*. This is also supported by the small subset of known genes that we looked at (i.e., the *Kv1* genes), where we manually confirmed the genes are not overly predicted in the *C. compressirostris* genome annotation. It is possible that the larger gene number in *C.*

*compressirostris* (relative to the other mormyrid *P. kingsleyae*) reflects true differences among these species. Our improved assembly with longer and fewer contigs may also have facilitated annotation of more genes.

### **KCNA genes in *C. compressirostris* genome**

It is generally accepted that there were two rounds of genome duplication in early vertebrate evolution, and an additional genome duplication event in the ancestor of teleost fishes [39]. These ancient duplications could be reflected in the gene tree of the *Kv1* gene family. The monophyletic clusters of (1) *KCNA5b* and *KCNA10a/b*, (2) *KCNA1a/b* and *KCNA2a/b*, and (3) *KCNA3a/b* and *KCNA6a/b* are indeed compatible with a scenario of three subsequent whole genome duplications (Fig. 6). *KCNA7a/b* genes were also identified from our genome. These genes were the only ones containing an intron, while all other *KCNA* genes are intronless. Our manual curation showed that the *KCNA7a* gene has two gene copies, which are profoundly diverged in one of its exons. They are otherwise very similar and are situated close to each other in the same contig, pointing towards a recent lineage-specific tandem duplication. The *KCNA5a* was not found in the genome. This could reflect the incompleteness of our genome, but this copy could also have been lost during evolution [39]. Gene loss in this gene family is not uncommon among teleost fish, and in some lineages such as zebrafish, pufferfish and medaka only four monophyletic clusters in the *Kv1* gene family were found [39].

*Kv1* genes are hypothesized to be potentially involved in the diversification of the EOD signal among mormyrid weakly electric fish. 13 *Kv1* genes are upregulated in the EO in the species *C. tshokwe* (a species with an elongated EOD) compared with skeleton muscle and *C. compressirostris* (a species with short EOD; [19]). While the *KCNA7b* gene is considered to be



upregulated in the skeleton muscle in mormyrids, the *KCNA7a* gene had is predominantly expressed in the EO [42]. This points towards one of the duplicated gene copies having evolved a new function (neofunctionalization) [43]. This might have occurred in mormyrid fishes, leading to more diverse functions among *Kv1* genes. In particular, the evolution of an electric organ may have exerted different selection pressures on ion channels, such that one paralog may have evolved a new function (in the EO), while the other maintains the original state. This could have fostered the retention of many *KCNA* genes, in comparison to other non-electric teleosts. Here, it is particularly interesting that we found *Campylomormyrus* to possess an additional copy of the *KCNA7a* gene. Not only is this gene known to be predominantly expressed in the electric organ, but expressed sequence differences in this gene have also been discussed underlying length modulation of the EOD [42]. Indeed, EOD divergence is considered a major driver of the radiation within the genus *Campylomormyrus* [18]. We found the exon 2 of the two *KCNA7a* duplicates exhibiting numerous expressed sequence variations. This exon encodes for mediating the voltage-dependent potassium ion permeability of excitable membranes, and the possession of two putatively functional copies may hence have facilitated divergent EOD evolution in *Campylomormyrus*. This hypothesis, however, still awaits evaluation by functional studies.

### **Gene family expansion and contraction in teleost**

The teleost-specific whole genome duplication has shaped the evolutionary history of many teleost lineages by providing extensive raw materials for species radiation [6]. A likely fate of many duplicated genes is also that they can become non-functional [7] as a result of lacking the selective constraint on preserving both genes. This may explain the global pattern of more contracted than expanded gene families in most teleost species. This pattern is only reversed in the two species representing Salmonidae and Cyprinidae, both having experienced an additional WGD [44, 45].

According to our CAFE5 analysis, Osteoglossomorpha appear to have a more rapid gene family turnover rate ( $\lambda$ ) than Otomorpha and Euteleostomorpha (Fig. 2&3). In particular, we found a significantly higher expansion/contraction ratio in Osteoglossomorpha, relative to other teleost lineages. This is exemplified by the *Kv1* gene family. Eight *Kv1* gene clusters (in total of 16 genes) were curated in *C. compressirostris* and an additional gene duplication detected (duplicating *KCNA7a*), while in other species such as the pufferfish, medaka, stickleback and zebrafish there are only four clusters. Although this is only a single gene family, it suggests a possible scenario of subfunctionalization and neofunctionalization in particular in the lineages with an active electric sense, which may contribute to the higher turnover rate and expansion/contraction ratio in Osteoglossomorpha.

### **Pathway evolution in Mormyroidea**

Pathway enrichment analysis is a tool to infer biologically relevant genes and biological processes from high-throughput data. The pathway of primary bile acid biosynthesis was most prone to gene family contraction in African weakly electric fish (mormyrids). This pathway takes place in the liver of vertebrates [46], where the synthesized bile acid can be conjugate with taurine or glycine before secretion via bile into the intestine. The pathway with most gene family expansion among mormyrids is nitrogen metabolism, one of the pathways for forming nitrogenous endproducts from protein degradation [47]. The expanded gene families contained within this second pathway are mostly related to carbonic anhydrases (e.g. CA12, CA4). These genes help maintaining acid-base homeostasis, regulating PH, and perhaps most relevantly, they play an active role in ion uptake [48]. It has been shown in mammals that genes such as carbonic anhydrases CA2 and CA4 play important roles in epithelial acid secretion and sodium uptake [48]. Although we do not know the

expression pattern of those CA genes in mormyrids, they might be involved in ion transport as well, especially of potassium and sodium, which are key to generate electric signals.

Expanded specifically in the electric mormyrids were gene families of the Wnt signaling pathway, encoding for a wide array of cellular processes including cell fate determination, motility, polarity, primary axis formation and organogenesis. It can be divided into the Planar Cell Polarity pathway and the Wnt/Ca<sup>2+</sup> pathway. High turnover was also observed in the calcium signaling pathway, which is mostly contracted in both species and their common ancestor. It regulates the Ca<sup>2+</sup> entering the cell from the outside. It was found to be down-regulated in the EO compared with skeleton muscle [21], which may have resulted from the contracted OGs about this pathway.

## **Conclusions**

A new high-quality genome of an African weakly electric fish (*C. compressirostris*, Mormyridae) is reported here, representing an important contribution to understand the evolution of electric fish and Osteoglossomorpha fish genomes. Our gene family analysis relative to representatives of many teleost fish genomes reveals a more rapid turnover rate and a higher expanded/contracted gene family number ratio in Osteoglossomorpha. The functional importance of these gene families requires further investigation, but provides many avenues for understanding the unique adaptations in these fishes. We also identified most of the *KCNA* gene clusters in our genome except for *KCNA5a*. The *KCNA7a* gene was found to be tandem duplicated. *KCNA* genes are considered of prime importance in the evolution of the active electric sense in teleosts. Our exhaustive efforts to localize these genes (including detection of a novel tandem duplication) underline the potential our new genome may hold towards an improved understanding of electric fish and Osteoglossomorpha evolution.

## **Methods**

### **Samples**

Genomic DNA was isolated from available frozen fin clips, which had been previously taken in the course of another study from an adult *C. compressirostris* artificially bred and raised at University Potsdam, Germany. The CTAB protocol was used to obtain high molecular weight genomic DNA [49]. The concentration and quality were further verified with Nanodrop spectrophotometer and Agilent TapeStation before sequencing.

### **Genome sequencing**

For Pacbio sequencing, a 15-kb SMRT cell DNA library was prepared and sequenced on a PacBio Sequel platform with one SMRT cell by a commercial company (Novogene). This produced 294 Gb long reads, which were used to generate the HiFi long reads using circular consensus sequencing (CCS) mode (Pacific Biosciences, USA).

### **De novo genome assembly**

The genome size and heterozygosity was estimated by GenomeScope 2.0 [50] using a k-mer value of 32 [28]. The genome was further assembled by hifiasm [29] with the HiFi reads as input. The separated primary haplotigs were visualized in Bandage [51]. This showed that some of the contigs contains two forks, which are likely homozygous breakpoints. Therefore, the program purge\_dups [52] was additionally applied for haplotig purging in the primary haplotigs. The mitochondrial DNA was separately assembled with the MitoHiFi.

We examined potential contamination using Blobtools2 [53] based on divergence in GC-content and coverage. We further assessed the presence of core, single copy and orthologous genes through BUSCO 5.3 [30] with the actinopterygii\_odb10 orthologues as reference.

### **Genome annotation**

Before annotation, we performed repeat masking in RepeatModeler 1.0.11 [31] provided in GenSAS v6.0 [54]. The soft repeat-masked sequence was used as an input in MAKER [55]. To provide EST evidence, we assembled the transcript sequences from Lamanna et al. [21] and newly generated RNA sequence data (Feng & Tiedemann, unpubl. results) of *C. compressirostris* with Trinity [56]. 575,330 transcripts were assembled by Trinity from electric organ and skeleton muscle tissues, and they were all used as EST evidence in MAKER. In addition, 24,4298 protein sequences were collected from all vertebrate proteins in NCBI.

The soft-masked assembly was predicted in MAKER with different gene predictors in three steps. In the first round, the RNA and protein sequences were supplied as evidence, and trained with the ab initio gene predictors SNAP [57] and Augustus [58] based on BUSCO. In the second round, we created a new SNAP-HMM input file based on the first round output and repeated the run with the same parameters as in the first round. The output from the second run was further analyzed in the third round following the method in the second round. The final output of the predicted gene, exon and intron information was statistically summarized by GAG [59]. We also manually checked the *KCNA* genes (see below) to exemplary confirm annotation quality. All the CDS were used to identify conserved protein domains with InterProScan [60].

### **Gene family expansion and contraction analysis**

In order to get an insight to the evolutionary dynamics of the genome evolution, the gene family analyzer CAFE5 [32] was used to infer expansion and contraction of gene ortholog clusters in 33 teleost fish. The species were selected such that they represent the taxonomic diversity among teleosts and include the only three available species from Osteoglossomorpha, six species from Otomorpha and 24 species from Euteleosteomorpha. A further selection criterion was genome quality, i.e. all selected representative species have a genome contig N50 over 100kb, except for that of *P. kingsleyae*, which was though retained, as it comprises the only other genome from an African electric fish (Mormyridae). We obtained the peptide sequences from those genomes in NCBI, and retained the longest isoform for each peptide. Gene families (orthogroups) were clustered among the filtered peptide sequences from all selected species in OrthoFinder [33], using an all-vs-all BLAST [38] for sequence similarity searches. Gene gain and loss in each lineage were calculated in CAFE5 with a random birth-death process model, based on the ultrametric species tree, which was generated by OrthoFinder using Fastree [61]. Taxon-specific lambda values (rates of evolutionary change) were estimated for Osteoglossomorpha, Otomorpha and Euteleosteomorpha.

In order to compare the gene family expansions and contractions relative to the most recent common ancestor (MRCA) of all selected teleost fishes, we summarized the gene and gene family change for each species and counted the ratios of expanded/contracted gene family (and gained/lost gene) numbers. A t-test was used to testify the ratios of expanded/contracted gene family numbers between Osteoglossomorpha & Euteleosteomorpha and Osteoglossomorpha & Otomorpha. Note that we only performed these two pairwise comparisons. The testing scheme is hence orthogonal and does not require a further correction [62].

We calculated the total number of orthogroups (OGs) from OrthoFinder and the number of significantly expanded and contracted OGs (P-value less than 0.05) among all species from the CAFE5 analysis. These OGs were separately blasted against the UniProt database. We further counted the OG number that showed up the most (zinc finger, transposon, immunoglobulin and GTPase) in both datasets, i.e. among all the OGs and the significantly contracted/expanded OGs. A Fisher's exact test [34] was then applied to identify significant deviations in the number of contracted/expanded OGs of each functional category, relative to the numbers of contracted/expanded OGs among all annotated OGs.

We collected the OGs inferred from electric fish and Osteoglossomorpha genomes (*C. compressirostris*, *P. kingsleyae*, *E. electricus* and *S. formosus*) and used the program VennDiagram [37] to visualize shared/unique OGs number among those four species. For *C. compressirostris*, *P. kingsleyae* and the ancestor node (as representative of Mormyridae), we performed an enrichment analysis by assigning contracted and expanded OGs to metabolic pathways using the KEGG database [36].

#### *KCNA* gene clusters curation

In total, we collected 233 *KCNA* genes sequences of teleost fishes in NCBI and blasted them against our genome. We identified each *KCNA* genes based on an e-value less than  $1e^{-6}$  and the best raw score from blast output. The identified *KCNA* genes were reciprocally blasted in NCBI. After we curated all found *KCNA* genes in our new *C. compressirostris* genome, a phylogenetic Bayesian tree was built in BEAST v1.8.4 [63] using a GTR+G substitution model, a relaxed lognormal clock model, the Yule speciation model and one billion MCMC. The result was

preserved only if the effective sample size (ESS) were all over 200. 10% of the starting MCMC was used as burn in and the remainder was used to generate a phylogenetic tree.

## **Declaration section**

### **Ethics approval and consent to participate**

The sample was taken from a fish specimen bred and kept at the University of Potsdam in compliance with German animal welfare regulations. Sampling followed the international recognized guidelines and applicable national law (Tierschutzgesetz). The procedure was approved by the deputy of animal welfare at University of Potsdam.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

The genome datasets generated during the current study are available in the European Nucleotide Archive under the accession number GCA\_910591475 at

[https://www.ebi.ac.uk/ena/browser/view/GCA\\_910591475.1](https://www.ebi.ac.uk/ena/browser/view/GCA_910591475.1).

The raw sequencing reads, assembled genome, annotation as well as the KCNA genes sequences were stored in Dryad under the DOI: doi:10.5061/dryad.c59zw3rcj. Prior to publication these data are available at:

[https://datadryad.org/stash/share/r9DNhYnIfU5xgBnt\\_\\_zKIQID60ieImk0Ti\\_vHDUS9n0](https://datadryad.org/stash/share/r9DNhYnIfU5xgBnt__zKIQID60ieImk0Ti_vHDUS9n0)

### **Competing interests**



The authors declare that they have no competing interests.

## **Funding**

This project was funded by the University of Potsdam.

## **Acknowledgement**

We thank Tonio Pieterek, M.Sc., who bred the fish in the aquarium. We also thank Thomas Inäbnit, M.Sc., for help during the assembly of the mitochondrial genome.

## **References**

1. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, et al. The medaka draft genome and insights into vertebrate genome evolution. *Nature*. 2007;447:714–9.
2. Volff JN. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb)*. 2005;94:280–94.
3. Jatllon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Maucell E, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*. 2004;431:946–57.
4. Gundappa MK, To TH, Grønvold L, Martin SAM, Lien S, Geist J, et al. Genome-Wide Reconstruction of Rediploidization Following Autopolyploidization across One Hundred Million Years of Salmonid Evolution. *Mol Biol Evol*. 2022;39:msab310.
5. Nakatani Y, Takeda H, Kohara Y, Morishita S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res*. 2007;17:1254–65.
6. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genom*. 2014;289:1045–60.
7. Hooper SD, Berg OG. On the nature of gene innovation: Duplication patterns in microbial genomes. *Mol Biol Evol*. 2003;20:945–54.
8. Stern DL. The genetic causes of convergent evolution. *Nat Rev Genet*. 2013;14:751–64.
9. Fund M, Phy NSF, Gallant JR, Traeger LL, Volkening JD, Moffett H, et al. Evolution of Electric Organs. *Science (1979)*. 2014;344:1522–5.

10. Wang Y, Yang L. Genomic Evidence for Convergent Molecular Adaptation in Electric Fishes. *Genome Biol Evol.* 2021;13:1–11.
11. Gallant JR, Traeger LL, Volkening JD, Moffett H, Chen PH, Novina CD, et al. Genomic basis for the convergent evolution of electric organs. *Science* (1979). 2014;344:1522–5.
12. Lissmann HW. On the Function and Evolution of Electric Organs in Fish. *J Exp Biol.* 1958;35:156–91.
13. Zakon HH, Zwickl DJ, Lu Y, Hillis DM. Molecular evolution of communication signals in electric fish. *J Exp Biol.* 2008;211:1814–8.
14. Crampton WGR. Electroreception, electrogenesis and electric signal evolution. *J Fish Biol.* 2019;95:92–134.
15. Glaubrecht M, Schneider H. Evolution in action: Case studies in adaptive radiation, speciation and the origin of biodiversity. *Evolution in Action: Case studies in Adaptive Radiation, Speciation and the Origin of Biodiversity.* 2010;:1–586.
16. Feulner PGD, Kirschbaum F, Mamonekene V, Ketmaier V, Tiedemann R. Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: *Campylomormyrus*): A combined molecular and morphological approach. *J Evol Biol.* 2007;20:403–14.
17. Denizot JP, Kirschbaum F, Westby GWM, Tsuji S. On the development of the adult electric organ in the mormyrid fish *Pollimyrus isidori* (with special focus on the innervation). *J Neurocytol.* 1982;11:913–34.
18. Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R. Magic trait electric organ discharge (EOD): Dual function of electric signals promotes speciation in African weakly electric fish. *Commun Integr Biol.* 2009;2:329–31.
19. Nagel R, Kirschbaum F, Tiedemann R. 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.* 2017;203:183–95.
20. Paul C, Kirschbaum F, Mamonekene V, Tiedemann R. Evidence for Non-neutral Evolution in a Sodium Channel Gene in African Weakly Electric Fish (*Campylomormyrus*, Mormyridae). *J Mol Evol.* 2016;83:61–77.
21. Lamanna F, Kirschbaum F, Waurick I, Dieterich C, Tiedemann R. Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). *BMC Genom.* 2015;16:1–17.
22. Lamanna F, Kirschbaum F, Tiedemann R. 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.* 2014;14:1222–30.

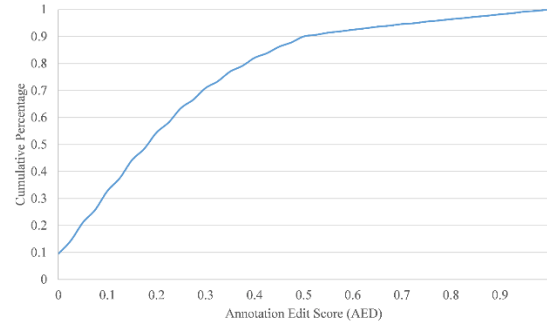
23. Feulner PGD, Kirschbaum F, Tiedemann R. Eighteen microsatellite loci for endemic African weakly electric fish (*Campylomormyrus*, Mormyridae) and their cross species applicability among related taxa. *Mol Ecol Notes*. 2005;5:446–8.
24. Arias CF, Dikow RB, McMillan WO, de León LF. De Novo Genome Assembly of the Electric Fish *Brachyhypopomus occidentalis* (Hypopomidae, Gymnotiformes). *Genome Biol Evol*. 2021;13:1–7.
25. Gallant JR, Losilla M, Tomlinson C, Warren WC. The genome and adult somatic transcriptome of the mormyrid electric fish *paramormyrops kingsleyae*. *Genome Biol Evol*. 2017;9:3525–30.
26. Traeger LL, Volkening JD, Moffett H, Gallant JR, Chen PH, Novina CD, et al. Unique patterns of transcript and miRNA expression in the South American strong voltage electric eel (*Electrophorus electricus*). *BMC Genom*. 2015;16.
27. Feulner PGD, Kirschbaum F, Tiedemann R. Adaptive radiation in the Congo River: An ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; *Campylomormyrus*). *J Physiol Paris*. 2008;102:340–6.
28. Marçais G, Kingsford C. Jellyfish : A fast k-mer counter. *Tutorialis e Manuais*. 2012;:1–8.
29. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18:170–5.
30. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol*. 2021;38:4647–54.
31. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117:9451–7.
32. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. *Bioinform*. 2020;36:5516–8.
33. Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14.
34. Upton GJG. Fisher ' s Exact Test. *Journal of the Royal Statistical Society*. 1992;155:395–402.
35. Kim TK. T test as a parametric statistic. *Korean J Anesthesiol*. 2015;68:540–6.
36. Yi Y, Fang Y, Wu K, Liu Y, Zhang W. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28:27–30.
37. Hanbo Chen PCB. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform*. 2011;12:1–7.
38. Ye J, McGinnis S, Madden TL. BLAST: Improvements for better sequence analysis. *Nucleic Acids Res*. 2006;34 WEB. SERV. ISS.:6–9.

39. Hoegg S, Meyer A. Phylogenomic analyses of KCNA gene clusters in vertebrates: Why do gene clusters stay intact? *BMC Evol Biol.* 2007;7:1–12.
40. Li J, Bian C, Hu Y, Mu X, Shen X. Data Descriptor : A chromosome- level genome assembly of the Asian arowana , *Scleropages formosus*. *Sci Rep.* 2016;105:1–8.
41. Delcher A L., Steven S L, Phillippy A M. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics.* 2003;1:10-3.
42. Swapna I, Ghezzi A, York JM, Markham MR, Halling DB, Lu Y, et al. Electrostatic Tuning of a Potassium Channel in Electric Fish. *Curr Biol.* 2018;28:2094–102.
43. Liu S. Evolution of Duplicate Gene Sequences , Expression Patterns , and Functions in the Brassicaceae and Other Rosids. University of British Columbia; 2011.
44. Xu P, Xu J, Liu G, Chen L, Zhou Z, Peng W, et al. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nat Commun.* 2019;10:1–11.
45. Kodama M, Briec MSO, Devlin RH, Hard JJ, Naish KA. Comparative mapping between coho salmon (*Oncorhynchus kisutch*) and three other salmonids suggests a role for chromosomal rearrangements in the retention of duplicated regions following a whole genome duplication event. *G3-GENES GENOM GENET.* 2014;4:1717–30.
46. Russell DW, Setchell KDR. Bile Acid Biosynthesis. *Biochem.* 1992;31:4737–49.
47. Patrick J. Walsh TPM. Evolutionary considerations of nitrogen metabolism and excretion. *Trends Endocrinol Metab.* 1996;7:153–4.
48. Hiroi J, McCormick S D. New insights into gill ionocyte and ion transporter function in euryhaline and diadromous fish. *RESPNB,* 2012, 184(3): 257-268.
49. Anghong P, Uengwetwanit T, Pootakham W, Sittikankaew K, Sonthirod C, Sangsrakru D, et al. Optimization of high molecular weight DNA extraction methods in shrimp for a long-read sequencing platform. *PeerJ.* 2020;8:1–18.
50. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinform.* 2017;33:2202–4.
51. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of de novo genome assemblies. *Bioinform.* 2015;31:3350–2.
52. Guan D, Guan D, McCarthy SA, Wood J, Howe K, Wang Y, et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 2020;36:2896–8.
53. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit - interactive quality assessment of genome assemblies. *G3-GENES GENOM GENET.* 2020;10:1361–74.

54. Lee T, Peace C, Jung S, Zheng P, Main D, Cho I. GenSAS - An online integrated genome sequence annotation pipeline. In: Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011. 2011. p. 1967–73.
55. Holt C, Yandell M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* 2011;12.
56. Grabherr MG, Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. N, Friedman and AR. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* 2013;29:644–52.
57. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinform.* 2008;24:2938–9.
58. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006;34:435–9.
59. Geib SM, Hall B, Derego T, Bremer FT, Cannoles K, Sim SB. Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. *Gigascience.* 2018;7:1–5.
60. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. *Bioinform.* 2014;30:1236–40.
61. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5.
62. Sokal RR, Rohlf FJ. *Biometry. The principles and practice of statistics in biological research.* 2012.
63. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:1–8.

## Additional files

### Additional file 1:



Annotation edit distance (AED) score distributions for the *C. compressirostris* annotation by MAKER. (PNG 168kb)

### Additional file 2:

Super Order	Order	Family	Species	Genome Size (mb)	Protein-coding Genes	Contig N50 (kb)	Reference/accession nr.
Osteoglossomorpha	Osteoglossiformes	Osteoglossidae	<i>Scleropages formosus</i>	784.5	23,537	9,102	GCF_900964775.1
Osteoglossomorpha	Osteoglossiformes	Mormyridae	<i>Campylomormyrus compressirostris</i>	862	34,492	1,399	this study; ERP129544
Osteoglossomorpha	Osteoglossiformes	Mormyridae	<i>Paramormyrops kingsleyae</i>	880	23,862	37.6	GCF_002872115.1
Otomorpha	Clupeiformes	Denticipitidae	<i>Denticiceps clupeioides</i>	567.4	23,488	3,060	GCA_900700375.2
Otomorpha	Cypriniformes	Danioiidae	<i>Danio rerio</i>	1,373	26,522	1,422	GCF_000002035.6
Otomorpha	Cypriniformes	Cyprinidae	<i>Cyprinus carpio</i>	1,680.1	43,531	1,559	GCF_018340385.1
Otomorpha	Gonorynchiformes	Chanidae	<i>Chanos chanos</i>	656.9	23,173	23,134	GCF_902362185.1
Otomorpha	Gonorynchiformes	Gymnotidae	<i>Electrophorus electricus</i>	589.4	22,304	104	GCA_013358815.1
Otomorpha	Characiformes	Serrasalminidae	<i>Pygocentrus nattereri</i>	1,222.1	25,548	12,899	GCA_015220715.1
Euteleosteiomorpha	Anabantiformes	Anabantidae	<i>Anabas testudineus</i>	555.6	23,850	7,055.436	GCA_900324465.3
Euteleosteiomorpha	Atheriniformes	Melanotaeniidae	<i>Melanotaenia boesemani</i>	865.6	24,098	9,299.978	GCA_017639745.1
Euteleosteiomorpha	Batrachoidiformes	Batrachoididae	<i>Thalassophryne amazonica</i>	2,446.6	22,351	2,329.598	GCA_902500255.1
Euteleosteiomorpha	Blenniiformes	Blenniidae	<i>Salaria fasciatus</i>	797.5	24,392	2,597.836	GCA_902148845.1
Euteleosteiomorpha	Carangiformes	Echeneidae	<i>Echeneis naucrates</i>	544.2	21,288	12,371.513	GCA_900963305.2
Euteleosteiomorpha	Centrarchiformes	Centrarchidae	<i>Micropterus salmoides</i>	963.6	27,179	1,227.323	GCA_014851395.1
Euteleosteiomorpha	Chaetodontiformes	Chaetodontidae	<i>Chelmon rostratus</i>	644.2	22,040	16,959.746	GCA_017976325.1
Euteleosteiomorpha	Cichliformes	Cichlidae	<i>Oreochromis aureus</i>	1,005.6	27,686	4,182.292	GCA_013358895.1
Euteleosteiomorpha	Cyprinodontiformes	Cyprinodontidae	<i>Cyprinodon tularosa</i>	1,086.9	23,979	1,362.15	GCA_016077235.1
Euteleosteiomorpha	Cyprinodontiformes	Poeciliidae	<i>Xiphophorus couchianus</i>	688.5	22,784	15,315.838	GCA_001444195.3
Euteleosteiomorpha	Esociformes	Esocidae	<i>Esox lucius</i>	918.7	24,647	22,630	GCA_011004845.1
Euteleosteiomorpha	Gadiformes	Gadidae	<i>Gadus morhua</i>	669.9	23,485	1,016	GCA_902167405.1
Euteleosteiomorpha	Gobiiformes	Gobiidae	<i>Periophthalmus magnuspinnatus</i>	752.6	21,306	2,301.275	GCA_009829125.1
Euteleosteiomorpha	Holocentriiformes	Holocentridae	<i>Myripristis murdjan</i>	835.3	23,439	14,476	GCA_902150065.1
Euteleosteiomorpha	Istiophoriformes	Xiphiidae	<i>Xiphias gladius</i>	691.8	21,362	5,252.707	GCA_016859285.1
Euteleosteiomorpha	Kurtiformes	Apogonidae	<i>Sphaeramia orbicularis</i>	1,342.7	24,116	2,360.121	GCA_902148855.1
Euteleosteiomorpha	Labriformes	Labridae	<i>Cheilinus undulatus</i>	1,173.5	23,316	16,477.222	GCA_018320785.1
Euteleosteiomorpha	Perciformes	Sebastidae	<i>Sebastes umbrosus</i>	800.9	23,881	11,445.908	GCA_015220745.1
Euteleosteiomorpha	Perciformes	Channichthyidae	<i>Pseudochannichthys georgianus</i>	1,026.1	23,287	661.283	GCA_902827115.1
Euteleosteiomorpha	Pleuronectiformes	Scophthalmidae	<i>Scophthalmus maximus</i>	538.2	21,619	25,762	GCA_022379125.1
Euteleosteiomorpha	Salmoniformes	Salmonidae	<i>Oncorhynchus mykiss</i>	2,341.7	41,896	15,580	GCA_013265735.3
Euteleosteiomorpha	Scombriformes	Scombridae	<i>Thunnus maccoyii</i>	782.4	24,659	26,803.536	GCA_910596095.1
Euteleosteiomorpha	Spariformes	Sparidae	<i>Acanthopagrus latus</i>	685.1	23,786	14,880.455	GCA_904848185.1
Euteleosteiomorpha	Syngnathiformes	Syngnathidae	<i>Syngnathus acus</i>	324.3	19,551	11,959.915	GCA_901709675.2

Information on teleost genomes used in the CAFE5 analysis. (XLSX 13kb)

### Additional file 3:

Species	Group	Gene Family			Gene		
		Expansion	Contraction	Ratio of Expansion/Contraction	Gain	Loss	Ratio of Gain/Loss
<i>Scleropages formosus</i>	Osteoglossomorpha	2,253	3,018	0.74	2,763	3,390	0.82
<i>Paramormyrops kingsleyae</i>	Osteoglossomorpha	2,667	3,353	0.79	3,574	4,167	0.86
<i>Campylomormyrus compressirostris</i>	Osteoglossomorpha	3,360	3,589	0.93	5,680	5,004	1.14
<i>Denticeps clupeioides</i>	Otomorpha	1,821	3,156	0.57	2,222	3,483	0.64
<i>Chanos chanos</i>	Otomorpha	822	2,803	0.29	1,813	3,072	0.59
<i>Danio rerio</i>	Otomorpha	1,370	2,983	0.46	2,901	3,177	0.91
<i>Cyprinus carpio</i>	Otomorpha	12,830	1,676	7.66	17,607	1,754	10.04
<i>Electrophorus electricus</i>	Otomorpha	762	3,113	0.245	1,286	3,617	0.36
<i>Pygocentrus nattereri</i>	Otomorpha	1,067	2,235	0.48	2,364	2,313	1.02
<i>Esox lucius</i>	Euteleosteomorpha	1,321	2,848	0.46	2,191	3,191	0.69
<i>Oncorhynchus mykiss</i>	Euteleosteomorpha	11,678	1,858	6.29	15,430	2,025	7.62
<i>Gadus morhua</i>	Euteleosteomorpha	889	4,096	0.22	2,399	4,468	0.54
<i>Myripristis murdjan</i>	Euteleosteomorpha	868	3,176	0.27	1,796	3,342	0.54
<i>Thalassophryne amazonica</i>	Euteleosteomorpha	1,296	4,253	0.30	1,797	4,778	0.38
<i>Syngnathus acus</i>	Euteleosteomorpha	653	5,161	0.13	1,012	5,389	0.19
<i>Sphaeramia orbicularis</i>	Euteleosteomorpha	1,556	3,531	0.44	2,482	3,703	0.67
<i>Periophthalmus magnuspinnatus</i>	Euteleosteomorpha	830	4,331	0.19	1,235	4,717	0.26
<i>Thunnus maccoyii</i>	Euteleosteomorpha	981	2,991	0.33	2,440	3,148	0.78
<i>Sebastes umbrosus</i>	Euteleosteomorpha	1,088	3,351	0.32	2,001	3,572	0.56
<i>Pseudochaenichthys georgianus</i>	Euteleosteomorpha	1,374	4,335	0.32	2,538	4,715	0.54
<i>Cheilinus undulatus</i>	Euteleosteomorpha	1,004	3,661	0.27	1,756	4,062	0.43
<i>Chelmon rostratus</i>	Euteleosteomorpha	889	3,496	0.25	1,313	3,955	0.33
<i>Acanthopagrus latus</i>	Euteleosteomorpha	1,043	3,148	0.33	1,814	3,338	0.54
<i>Micropterus salmoides</i>	Euteleosteomorpha	2,649	3,105	0.85	4,375	3,268	1.34
<i>Anabas testudineus</i>	Euteleosteomorpha	1,076	3,367	0.32	2,218	3,599	0.62
<i>Xiphias gladius</i>	Euteleosteomorpha	715	3,660	0.20	1,033	4,216	0.25
<i>Echeneis naucrates</i>	Euteleosteomorpha	816	4,020	0.20	1,290	4,476	0.29
<i>Scophthalmus maximus</i>	Euteleosteomorpha	771	3,690	0.21	1,071	4,257	0.25
<i>Salarias fasciatus</i>	Euteleosteomorpha	2,173	3,961	0.55	3,407	4,358	0.78
<i>Oreochromis aureus</i>	Euteleosteomorpha	1,552	3,145	0.49	4,342	3,302	1.31
<i>Melanotaenia boesemani</i>	Euteleosteomorpha	1,184	3,454	0.34	2,463	3,699	0.67
<i>Cyprinodon tularosa</i>	Euteleosteomorpha	1,340	3,724	0.36	2,409	3,989	0.60
<i>Xiphophorus couchianus</i>	Euteleosteomorpha	1,128	3,931	0.29	1,844	4,322	0.43

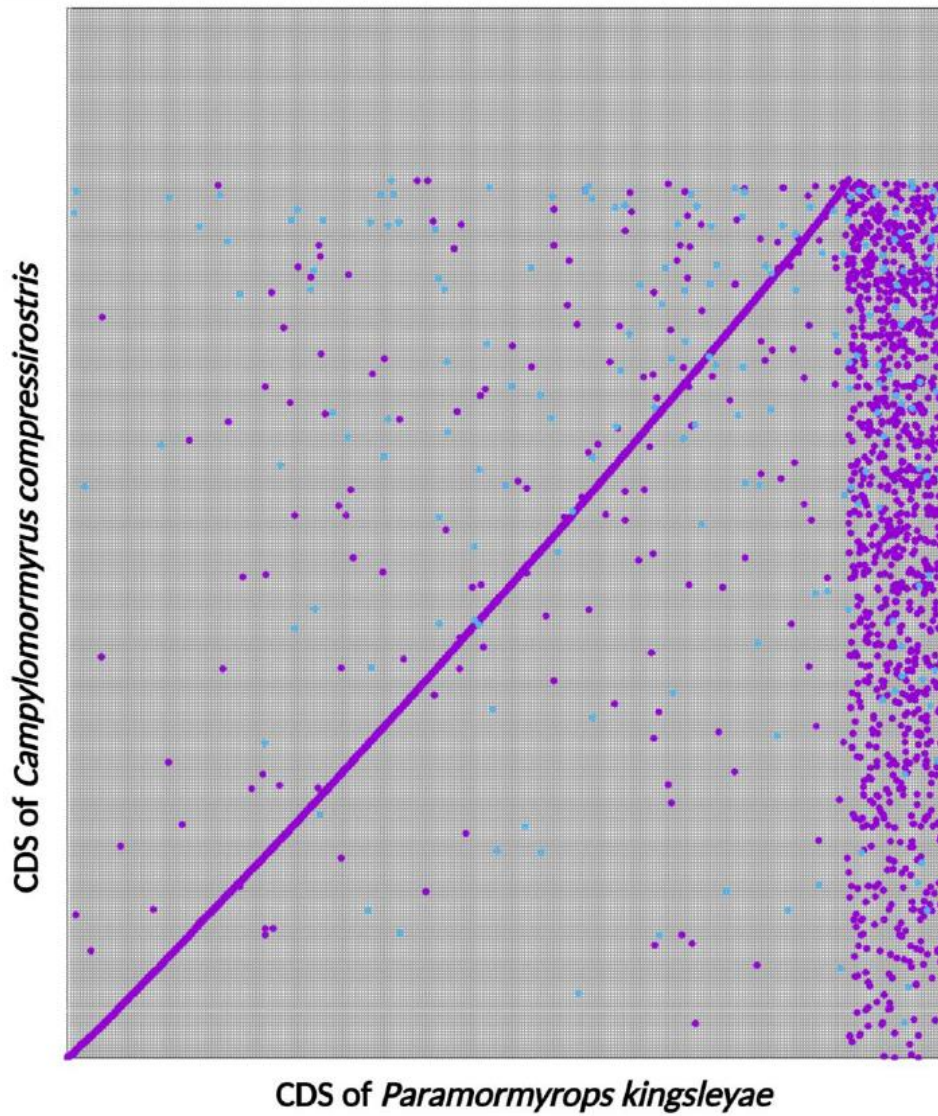
Number of gene and gene family expansions and contractions compared with MRCA, and ratio of the expanded/contracted gene family (gain/loss gene) numbers. (XLSX 12kb)

### Additional file 4:

Please find this file in the link: <https://link.springer.com/article/10.1186/s12864-023-09196-6#Sec22>

KCNA gene sequences. (TXT 25kb)

Additional file 5:



MUMmer alignment between the CDS of *C. compressirostris* and *P. kingsleyae*. (PNG 2.33mb)



## 4 Article II

### **Gene and allele specific expression underlying the electric signal divergence in African weakly electric fish.**

Feng Cheng<sup>1</sup>, Alice B. Dennis<sup>1, 2</sup>, Otto Baumann<sup>3</sup>, Frank Kirschbaum<sup>1, 4</sup>, Salim Abdelilah-Seyfried<sup>3</sup>, Ralph Tiedemann<sup>1,\*</sup>

1 Unit of Evolutionary Biology and Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

2 Laboratory of Adaptive Evolution and Genomics, Research Unit of Environmental and Evolutionary Biology, Institute of Life, Earth & Environment, University of Namur, Namur, Belgium

3 Department of Animal Physiology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

4 Department of Crop and Animal Science, Faculty of Life Sciences, Humboldt University, Berlin, Germany

**In the African weakly electric fish genus *Campylomormyrus*, electric organ discharge (EOD) signals are strikingly different in shape and duration among closely related species, contribute to pre-zygotic isolation and may have triggered an adaptive radiation. We performed mRNA sequencing on electric organs (EOs) and skeletal muscles (SMs; from which the EOs derive) from three species with short (0.4 ms), medium (5 ms), and long (40 ms) EODs and two different cross-species hybrids. We identified 1,444 up-regulated genes in EO shared by all five species/hybrids cohorts, rendering them candidate genes for EO-specific properties in *Campylomormyrus*. We further identified several candidate genes,**

**including *KCNJ2* and *KLF5*, their up-regulation may contribute to increased EOD duration. Hybrids between a short (*C. compressirostris*) and a long (*C. rhynchophorus*) discharging species exhibit EODs of intermediate duration and showed imbalanced expression of *KCNJ2* alleles, pointing towards a cis-regulatory difference at this locus, relative to EOD duration. *KLF5* is a transcription factor potentially balancing potassium channel gene expression, a crucial process for the formation of an EOD. Unraveling the genetic basis of the species-specific modulation of the EOD in *Campylomormyrus* is crucial for understanding the adaptive radiation of this emerging model taxon of ecological (perhaps even sympatric) speciation.**

Electric fish have independently evolved six times<sup>1-3</sup>. They possess a specific myogenic electric organ (EO) derived from skeletal muscle (SM) fibers except for Apterontidae which possess an EO derived from nervous tissue<sup>4</sup>. Comparative genomics have unraveled this convergent phenotypic evolution to originate in part also from convergence on the molecular level: both voltage-dependent sodium and potassium channels are involved in the electric organ development and physiology. Because of the teleost-specific whole genome duplication<sup>5</sup>, these fish possess two copies of most genes and subfunctionalization among paralogs and differential expression between EO and SM seem to play a major role in the transition of myocytes to electrocytes. A prominent example is the voltage-gated sodium ( $Na_v$ ) channel gene (*SCN4a*): convergently in three electrogenic taxa (Mormyroidea, Siluriformes and Gymnotiformes), only one paralog (*SCN4ab*) is still expressed in SM, but the other one (*SCN4aa*) is exclusively expressed in the EO, indicating a crucial role for electrogenesis<sup>6-8</sup>. The  $Na_v$  channel (*SCN4aa*) is regulated by *FGF13a* in the three electric fish lineages Siluriformes, Gymnotiformes, and Mormyroidea<sup>9</sup>. Differential expression of

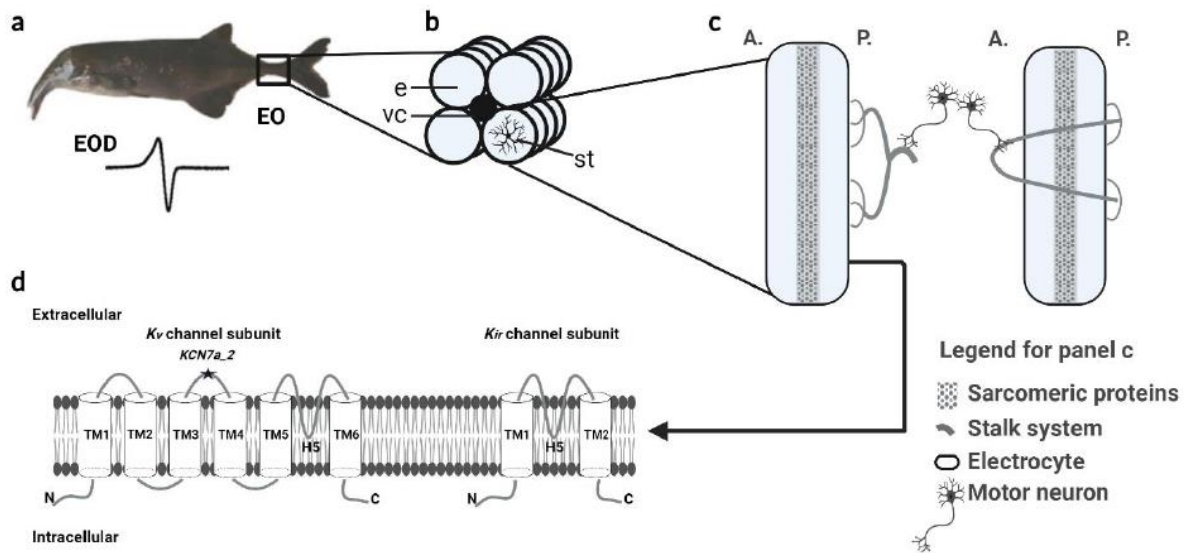
multiple isoforms of  $\alpha$  and  $\beta$  subunits of sodium/potassium ATPase is important in the EO as well<sup>9-11</sup>. In addition, several transcription factors, *HEY1*, *MEF2a*, *SIX2a*, are convergently up-regulated in the EOs of those electric fishes lineages<sup>9,10,12</sup>. EOs hence comprise a prime example of convergent evolution in both genotype and phenotype.

One of the electric fish clades, mormyrid fish, contains about 200 described species that are endemic to Africa. This outstanding adaptive radiation within the otherwise species-poor basal lineage of osteoglossiforms is putatively due to their species-specific weak electric signals, which is used for both electrolocation and electrocommunication<sup>13,14</sup>. Divergence in electric organ discharge (EOD) is considered a major driver in the ecological (and possibly sympatric) speciation in the mormyrid genus *Campylomormyrus*, which is mainly distributed in the Congo River<sup>15</sup>.

The genus *Campylomormyrus* comprises 15 described species, which have profoundly diverged in their electric organ discharge (EOD) with regard to signal duration and waveform<sup>13</sup>. Those species possess either long or short, biphasic or triphasic, but always species-specific EODs, that function as a pre-zygotic reproductive isolation mechanism and are supposed to have arisen via divergent selection among closely related species<sup>13</sup>. In adult *Campylomormyrus*, the electric organ, confined to the caudal peduncle (Fig. 1a), is composed of specialized electrocytes<sup>16</sup>. They have a flat, disk-shaped appearance with a clear orientation toward the longitudinal body axis (Fig. 1b). Unlike skeletal muscle myocytes, electrocytes possess a number of special evaginations, called stalks, mostly on the posterior face<sup>16</sup>. These stalks are either fused into major stalks on the posterior face (Fig. 1c left) or they penetrate the electrocyte and merge at the anterior face to constitute to major stalks (Fig. 1c right). A branch of the spinal nerve forms numerous synapses with the major stalk, whether on the posterior or on the anterior face of the electrocyte, and the action potentials are propagated along the stalk system to the disc-like part of the electrocyte<sup>16</sup>. The externally

measurable EOD is formed by simultaneous action potentials of all electrocytes. The shape of the EOD in *Campylomormyrus* is often associated with the penetration of the stalks<sup>17</sup>, while the structural basis of the EOD duration, which can vary 100-fold across species, is still only partially understood. A very elongated EOD (~40 ms) is produced by *C. rhynchophorus* and *C. numenius* which exhibit large foldings or evaginations on the anterior face of the electrocytes, so called papillae<sup>18,19</sup>. In two species with relatively short EOD (Fig. 2a), *C. compressirostris* (0.4 ms) and *C. tamandua* (0.4 ms), many small stalks fuse into one major stalk of large diameter after their origin<sup>16</sup>. In contrast, the stalk system in species with an EOD of medium (e.g. *C. tshokwe*, 5 ms), or long duration (e.g. *C. numenius*, 40 ms) is more branched<sup>16</sup>. Apart from these difference in the stalk system, species with highly diverged EOD waveforms still show similar electrocyte geometry suggesting further core mechanisms to contribute to the observed EOD variations. Since the electrocytes generate action potentials for EOD, the distribution and repertoire of ion currents have long been considered to play a key role in EOD formation<sup>11,20-23</sup>.

Sodium and potassium fluxes are considered the most important ion currents in controlling the EOD<sup>24</sup>. They are the basic requirements for generating an action potential<sup>25</sup>. Consequently, abundance and properties of sodium and potassium channels are likely to profoundly influence the EOD. The potassium channels can be classified into different classes based on their structure and function: voltage-gated ( $K_v$ , includes subfamilies e.g. *shaker*-related *KCNA*, *shab*-related *KCNB*), inwardly rectifying ( $K_{ir}$ ), tandem pore domain channels ( $K_{2p}$ ), ligand-gated channels and calcium-activated channels ( $K_{ca}$ )<sup>26</sup>. Two paralogs of the *KCNA7* channel gene originate from the whole genome duplication event in teleost fish and these paralogs might have undergone subfunctionalization or neofunctionalization in mormyrids: one of them *KCNA7a* is predominantly expressed in the EO of mormyrids, while *KCNA7b* is preferentially expressed in SM<sup>23</sup>. The  $K_v$

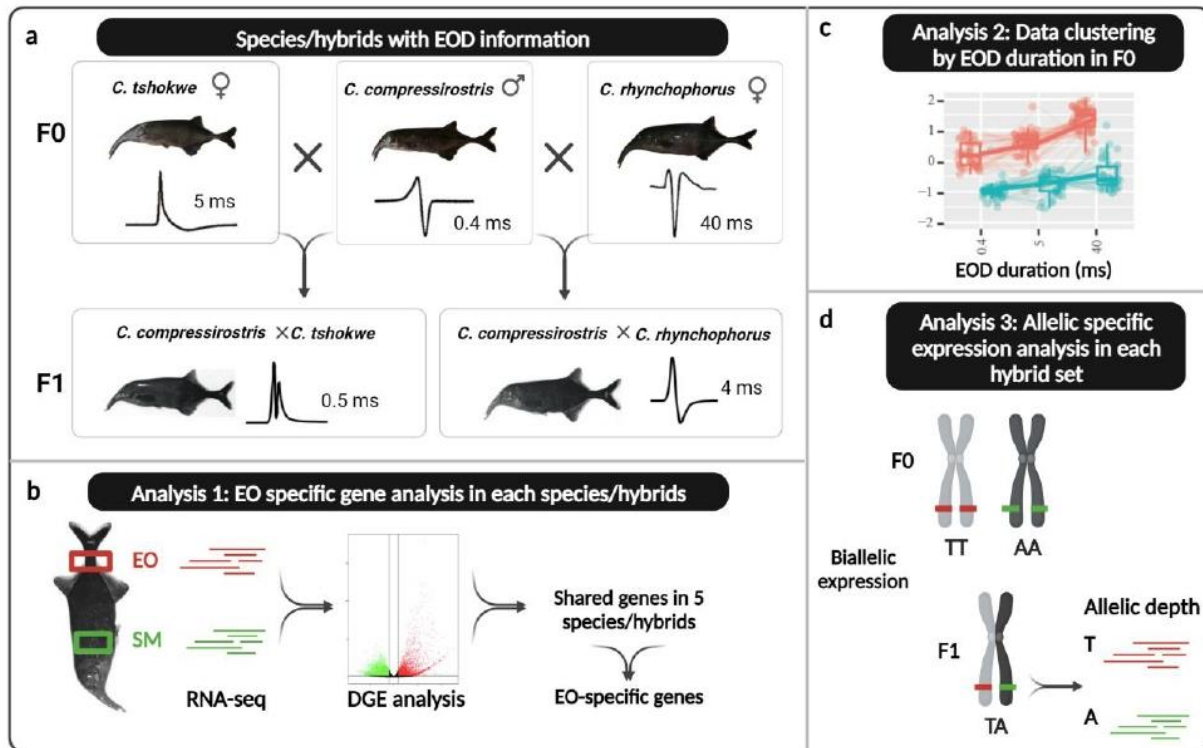


**Fig. 1** Electric organ and electrocyte structure in *Campylomormyrus*, and schematic illustration for potassium channels.

**a** Electric organ (EO) and electric organ discharge (EOD) in an adult *Campylomormyrus* fish. **b** The EO consists of four columns of electrocytes (e) which surround the vertebral column (vc), the stalk system (st) is connected to the posterior face of the electrocyte. **c** Anterior (A.) and posterior (P.) faces of electrocytes with two types of stalk system. Panel c is modified from Gallant *et al.* 2012. **d** Schematic illustration of voltage-gated potassium ( $K_v$ ) channel and inwardly rectifying ( $K_{ir}$ ) channel subunit.  $K_v$  channel subunit contains six transmembrane (TM) helices, a pore-forming (H5) loop, and cytosolic  $NH_2$  (N) and  $COOH$  (C) termini. The gene *KCN7A\_2* was inferred to be under positive selection and the mutation encodes the loop between TM3-4.  $K_{ir}$  channel subunit contains only two TMs.

channel contains six transmembrane helices (Fig. 1d). In *KCNA7a*, a non-synonymous substitution was observed in the transmembrane helices 3-4 linker and the encoded amino acid substitution might relate to the EOD duration difference among the mormyrid taxa *Brienomyrus* and *Gymnarchus*<sup>23</sup>.

This study focusses on potential molecular mechanisms underlying the divergent EOD among *Campylomormyrus* species as a potential major driver of their adaptive radiation. This study takes further advantage of artificially bred hybrid electric fish. *Campylomormyrus* species hybrids often exhibit an adult EOD which is similar to the juvenile EOD from one of the parental species, and the adult EOD duration in hybrids is usually intermediate between the two parental species<sup>19</sup>. Gene



**Fig. 2** Electric organ discharge (EOD) shape and duration of *Campylomormyrus* species and hybrids, and the working flow of this study.

**a** Species/hybrids samples used in the study and their EOD pattern. **b** Differential gene expression (DGE) analysis between electric organ (EO) and skeletal muscle (SM) for each species/hybrid to identify genes with EO-specific expression. **c** RNA-seq data clustering based on EOD duration change EO (red) and SM (blue) in F0 species. **d** Allele specific expression analysis in each hybrid set.

expression analyses in hybrids further enable assessment of allelic specific expression, relative to the expressed trait of interest (here, EOD duration). To enhance our understanding of the genetic regulation of EOD divergence among *Campylomormyrus* species, especially for the EOD duration divergence, we: 1) compared the gene expression pattern between electric organ (EO) and skeletal muscle (SM) in the three F0 species *C. compressirostris* (*com*, short and biphasic EOD), *C. tshokwe* (*tsh*, medium and biphasic EOD), *C. rhynchophorus* (*rhy*, long and triphasic EOD), and two F1 hybrids *C. compressirostris* ♂ × *C. tshokwe* ♀ (*com* × *tsh*, short and biphasic EOD), and *C. compressirostris* ♂ × *C. rhynchophorus* ♀ (*com* × *rhy*, medium and biphasic EOD); 2) clustered RNA-seq data relative to the EOD duration in three F0 species to infer genes with duration-specific

expression; 3) assessed biallelic specific expression for two hybrid sets (each set includes two F0 parental species and their hybrid; Fig. 2).

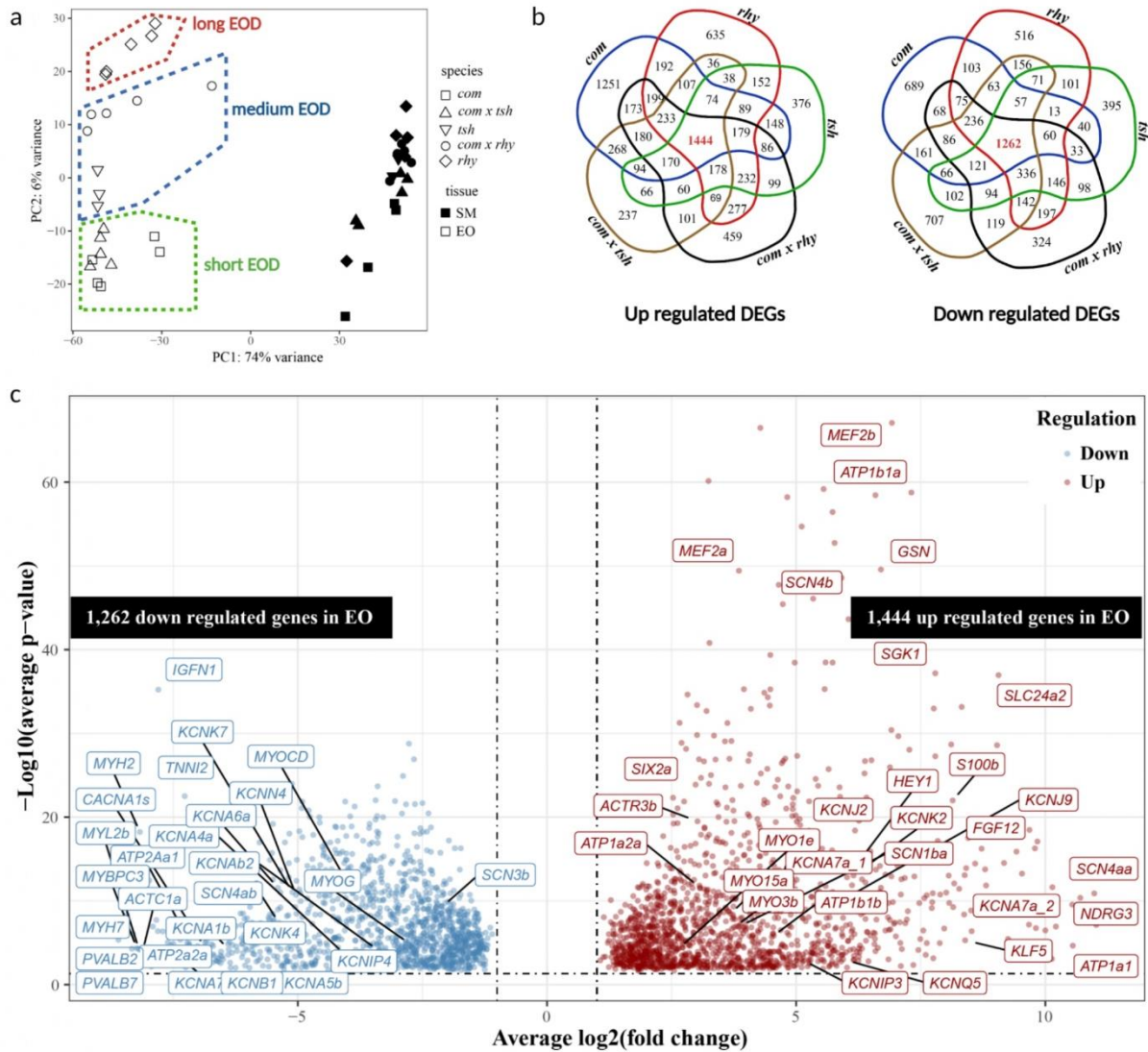
## Results

We examined overall patterns in gene expression using a principal component analysis (PCA) based on all expressed genes (Fig. 3a). Expression profiles of SMs and EOs were broadly separated along PC1, which explained 74% of variance. The SMs profiles from all species/hybrids clustered together; however, species/hybrid-specific EOs' expression profiles were stratified along PC2 (explained 6% of variance), relative to EOD duration (Fig. 3a). The PCA hence indicates that gene expression in *Campylomormyrus* is (1) EO specific, compared with SM; (2) relates to EOD duration, enabling inference of underlying candidate genes.

### Genes with EO-specific expression pattern

Differential gene expression analysis was used for pairwise comparisons between EO and SM for each species and hybrid (Fig. 2b). We identified significantly differentially expressed genes (DEGs) based on a  $|\log_2 \text{fold change} (\log_2 \text{FC})| > 1$  and a p-value  $< 0.05$ . We specifically identified genes with an EO-specific expression pattern shared among all *Campylomormyrus* species/hybrids. There were 1,444 up-regulated and 1,262 down-regulated DEGs that were shared in the comparison of EO and SM in all species/hybrids (Fig. 3b, c).

Among the DEGs up-regulated in the EO, 54 genes were related to transmembrane ion transport (Fig. 3c, Table 1, Supplementary Table 1). We identified four genes encoding sodium/potassium-ATPase  $\alpha$  and  $\beta$  subunit (*ATP1a1*, *ATP1a2a*, *ATP1b1a* and *ATP1b1b*), and three  $Na_v$  channel genes (*SCN4aa*, *SCN4b* and *SCN1ba*). Several genes encoding for different types of potassium channels were also identified: four  $K_v$  channel genes (*KCNA7a\_1*, *KCNA7a\_2*, *KCNIP3* and *KCNQ5*), two



**Fig. 3** Differential gene expression between electric organ (EO) and skeletal muscle (SM) in *C. compressirostris* (*com*), *C. rhynchophorus* (*rhy*), *C. tshokwe* (*tsh*) and hybrids *C. compressirostris* ♂ x *C. rhynchophorus* ♀ (*com x rhy*), *C. compressirostris* ♂ x *C. tshokwe* ♀ (*com x tsh*).

**a** Principal component analysis (PCA) of gene expression levels between EO and SM in 5 species/hybrids. **b** Venn Diagram graph for up (left) and down (right) regulated genes shared in 5 species/hybrids. All differentially expressed genes (DEGs) have  $|\log_2(\text{fold change})| > 1$  and a  $p\text{-value} < 0.05$ . Many of the DEGs are related to "membrane" and "plasma membrane" (see Supplementary Fig. 1). **c** Volcano plot showing genes differentially expressed in EO (relative to SM) in all 5 species/hybrids. X-axis is the average  $\log_2(\text{fold change})$  among 5 species/hybrids, and y-axis is the associated  $-\log_{10}(\text{average } p\text{-value})$  for 5 species/hybrids. Potential candidate genes and genes with low p-value or high fold change are labeled with their name.

*K<sub>ir</sub>* channel genes (*KCNJ2* and *KCNJ9*), and one *K<sub>2p</sub>* channel gene (*KCNK2*). Further transmembrane ion transport DEGs were chloride, calcium and other cation channel genes



(Supplementary Table 1). Several solute carrier family genes were also up-regulated in the EO, in particular *SLC24a2* (Table 1).

18 genes up-regulated in the EO were associated with cytoskeletal and sarcomeric protein (Supplementary Table 1). The predicted function of those genes were mainly related to F-actin dynamics and unconventional myosin activity (Table 1). A signaling gene *NDRG3* showed very high overexpression in EO ( $\log_2FC=11.02$ ), as well as the genes *SGK2*, *S100b* and *FGF12*. The up-regulated transcription factors in the EO included *KLF5*, *FOXL2*, *SIX2a*, *HEY1* and two myocyte-specific enhancer factors (*MEF2a* and *MEF2b*).

In the down-regulated DEGs in EO (or up-regulated in SM), 44 genes were classified into the category “cytoskeletal & sarcomeric” (Fig. 3c, Supplementary Table 2). There were 37 transmembrane ion transport genes down-regulated in EO, which were related to the ions potassium, sodium, and calcium. In contrast to the expression pattern of the two *KCNA7a* copies, five *K<sub>v</sub>I* subfamily genes (*KCNA1b*, *KCNA4a*, *KCNA5b*, *KCNA6a* and *KCNA7b*) were down-regulated in the EO. This was also the case for other potassium and sodium channel genes, e.g. *K<sub>v</sub>* subfamily genes (*KCNB1*, *KCNE4*, *KCNIP4*), *K<sub>2p</sub>* subfamily genes (*KCNK4*, *KCNK7*), a *K<sub>ca</sub>* subfamily gene (*KCNN4*), and *Na<sub>v</sub>* channel genes (*SCN3b*, *SCN4ab*). Two muscle-specific transcription factors, *MYOCD* and *MYOG*, were also down-regulated in EO (Supplementary Table 2).

We applied a Gene Ontology (GO) enrichment analysis to further examine the function of all the up- and down-regulated DEGs in EO respectively<sup>27</sup>. Among the up-regulated DEGs in the EO, there were 44 significantly enriched GO terms (Fisher’s exact test p-value<0.01, Supplementary Fig. 1, Supplementary

**Table 1** Candidate genes up-regulated in all species/hybrids in the electric organ relative to skeletal muscle.

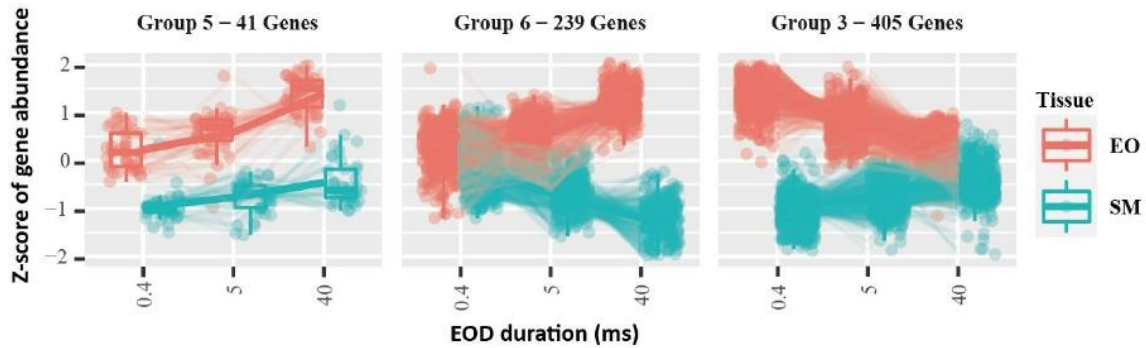
ID	Blast Gene	Highlights of Predicted Function	Gene Description	Category	Average log2FC	Average Pvalue
maker-ptg0003611-augustus-gene-0.2-mRNA-1	<i>ACTR3b</i>	F-actin dynamics / polymerization	ARP3 actin related protein 3 homolog B	cytoskeletal & sarcomeric	3.02	3.16E-20
maker-ptg0003461-snap-gene-25.173-mRNA-1	<i>GSN</i>	F-actin dynamics / polymerization	gelsolin	cytoskeletal & sarcomeric	6.70	2.7189E-50
snap_masked-ptg0000281-processed-gene-137.57-mRNA-1	<i>MYO15a</i>	unconventional myosin; actin-based motor protein	unconventional myosin-XV	cytoskeletal & sarcomeric	3.68	3.4002E-09
maker-ptg0010031-snap-gene-2.16-mRNA-1	<i>MYO1e</i>	unconventional myosin; actin-based motor protein	unconventional myosin-Ie	cytoskeletal & sarcomeric	2.63	5.7074E-05
maker-ptg0020901-augustus-gene-35.16-mRNA-1	<i>MYO3b</i>	unconventional myosin; actin-based motor protein	myosin-IIIb	cytoskeletal & sarcomeric	3.75	1.4271E-07
maker-ptg0002151-snap-gene-13.31-mRNA-1	<i>S100b</i>	cytosolic Ca <sup>2+</sup> -binding protein of the EF-hand superfamily	S100 calcium binding protein B	signaling	8.14	1.0122E-22
maker-ptg0000491-snap-gene-23.20-mRNA-1	<i>FGF12</i>	possibly regulate voltage-gated sodium channels	fibroblast growth factor 12	signaling	8.72	5.8498E-17
maker-ptg0008381-snap-gene-3.218-mRNA-1	<i>NDRG3</i>	predicted to be involved in signal transduction	N-myc downstream-regulated gene 3 protein	signaling	11.02	7.4446E-08
maker-ptg0015631-augustus-gene-5.35-mRNA-1	<i>SGK1</i>	serine/threonine-protein kinase	serine/threonine-protein kinase Sgk1	signaling	7.79	6.6327E-38
maker-ptg0010881-snap-gene-0.7-mRNA-1	<i>SIX2a</i>	target ARE promoter elements in sodium/potassium adenosine triphosphatases	SIX homeobox 2	transcription factor	3.05	1.221E-27
maker-ptg0007831-snap-gene-6.78-mRNA-1	<i>HEY1</i>	developing cardiac conduction pathway	hes related family bHLH transcription factor with YRPW motif 1	transcription factor	6.02	1.5325E-13
maker-ptg0017401-augustus-gene-1.112-mRNA-1	<i>KLF5</i>	rebalance potassium channels	Krüppel-like factor 5	transcription factor	8.39	5.0491E-06
maker-ptg0000081-snap-gene-10.43-mRNA-1	<i>MEF2a</i>	transcriptional activator for numerous muscle-specific genes	myocyte-specific enhancer factor 2A	transcription factor	3.85	3.7818E-50
maker-ptg0012701-snap-gene-47.16-mRNA-1	<i>MEF2b</i>	transcriptional activator for numerous muscle-specific genes	myocyte-specific enhancer factor 2B	transcription factor	6.92	8.4583E-68
maker-ptg0009701-augustus-gene-2.127-mRNA-1	<i>ATP1a1</i>	sodium/potassium-ATPase $\alpha$ -subunit	sodium/potassium-transporting ATPase subunit alpha-1	transmembrane ion transport	10.55	2.1894E-05
snap_masked-ptg0011561-processed-gene-0.19-mRNA-1	<i>ATP1a2a</i>	sodium/potassium-ATPase $\alpha$ -subunit	sodium/potassium-transporting ATPase subunit alpha-2	transmembrane ion transport	3.07	2.6317E-12
maker-ptg0010471-snap-gene-1.63-mRNA-1	<i>ATP1b1a</i>	sodium/potassium-ATPase $\beta$ -subunit	ATPase sodium/potassium transporting beta 1a	transmembrane ion transport	6.59	3.6325E-59
maker-ptg0005091-snap-gene-9.39-mRNA-1	<i>ATP1b1b</i>	sodium/potassium-ATPase $\beta$ -subunit	ATPase sodium/potassium transporting beta 1b	transmembrane ion transport	4.51	2.0647E-06
maker-ptg0000281-snap-gene-81.10-mRNA-1	<i>KCNA7a_1</i>	Kv channel	potassium voltage-gated channel subfamily A member 7a	transmembrane ion transport	4.60	2.5847E-11
maker-ptg0000281-snap-gene-81.8-mRNA-1	<i>KCNA7a_2</i>	Kv channel	potassium voltage-gated channel subfamily A member 7a	transmembrane ion transport	8.27	3.5553E-12
maker-ptg0014271-snap-gene-13.20-mRNA-1	<i>KCNIP3</i>	Kv channel	calnenilin	transmembrane ion transport	5.11	0.00099357
maker-ptg0006971-snap-gene-6.109-mRNA-1	<i>KCNQ5</i>	Kv channel	potassium voltage-gated channel subfamily Q member 5	transmembrane ion transport	5.94	0.00093894
maker-ptg0002651-est_gff_est2genome-gene-6.33-mRNA-1	<i>KCNJ2</i>	Kir channel	inward rectifier potassium channel 2	transmembrane ion transport	5.53	1.1222E-20
maker-ptg0008301-augustus-gene-5.123-mRNA-1	<i>KCNJ9</i>	Kir channel	G protein-activated inward rectifier potassium channel 3	transmembrane ion transport	5.77	2.6324E-10
snap_masked-ptg0011181-processed-gene-0.13-mRNA-1	<i>KCNK2</i>	K2p channel	potassium channel subfamily K member 2	transmembrane ion transport	6.15	9.238E-14
maker-ptg0002531-augustus-gene-20.10-mRNA-1	<i>SCN1ba</i>	Nav channel	sodium channel subunit beta-1	transmembrane ion transport	3.84	1.2983E-07
maker-ptg0011881-snap-gene-6.4-mRNA-1	<i>SCN4aa</i>	Nav channel	sodium channel protein type 4 subunit alpha A	transmembrane ion transport	10.98	1.2737E-11
maker-ptg0022391-snap-gene-5.5-mRNA-1	<i>SCN4b</i>	Nav channel	sodium voltage-gated channel beta subunit 4	transmembrane ion transport	5.34	8.1806E-47
maker-ptg0001481-snap-gene-9.4-mRNA-1	<i>SLC24a2</i>	calcium, potassium:sodium antiporter	solute carrier family 24 member 2	transmembrane ion transport	9.06	1.1057E-37

Table 3). Among them, the three GO terms with the highest number of DEGs were all related to the cell membrane: membrane (464 DEGs), integral component of membrane (309 DEGs) and plasma membrane (237 DEGs). There were 47 DEGs assigned to the enriched GO term “ion transport”. 62 and 35 DEGs were assigned to the enriched Golgi-related GO terms “Golgi membrane” and “Golgi apparatus”, respectively. In addition, there were 23 DEGs assigned to the enriched GO term “actin filament binding”. There were 73 GO terms significantly enriched for DEGs down-regulated in the EO (up-regulated in SM, Supplementary Fig. 2, Supplementary Table 4). They were associated with skeletal and cardiac muscle tissue related GO terms.

#### **Genes with expression levels related to EOD duration.**

The PCA plot from transcriptome-wide gene expression showed a striking association between overall gene expression and EOD duration in all species/hybrids (PC2 in Fig. 3a; accounting for 6% of the variance in gene expression). DESeq2 provides a Likelihood Ratio Test (LRT) that compares how well a gene’s read count data fit a “full model” (with independent variables) compared to a “reduced model” (without those variables). Therefore, it is well suited to explore whether there are any significant associations of gene expression levels across a series of values of an independent variable (here, EOD duration)<sup>28</sup>. Specifically, we used this approach to test whether a gene’s expression fits a pattern of increasing or decreasing over the different durations in two different tissues, EO and SM<sup>29</sup>. In order to avoid any bias potentially stemming from distorted expression pattern in the hybrids, we only used the quantification data from the parental pure-bred (F0) species. The LRT analysis returned 1,874 significant genes using a threshold of  $\text{padj} < 0.05$ . Those genes were further sorted into groups using the `degPatterns` function. Each such group contained genes following a specific pattern of expression across the different duration values in the analyzed tissues EO and SM<sup>30</sup>.

The degPatterns function generated 27 groups of different expression pattern in EO and SM, relative to EOD duration (Supplementary Fig. 3). To identify EOD duration-specific genes, we focused on the groups meeting the following criteria: 1) the gene expression level in EO is higher than SM in all F0 species (i.e., the gene is consistently up-regulated in the EO); 2) the gene expression level in the EO shows an increasing or decreasing pattern, relative to EOD duration.



**Fig. 4** RNA-seq data clustered by EOD duration (only for the 3 pure-bred species).

Increasing (**Group 5** and **6**) and decreasing (**Group 3**) expression patterns over EOD duration among electric organ (EO) and skeletal muscle (SM). The x-axis for each group represents the duration of the analyzed species: *C. compressirostris* (0.4ms), *C. tshokwe* (5ms) and *C. rhynchophorus* (40ms).

Two groups showed a consistent increasing expression pattern (groups 5, 41 genes; and 6, 239 genes) and one a decreasing expression pattern (group 3, 405 genes), relative to the EOD duration (Fig. 4).

In the increased expression pattern groups (5 and 6), we found  $K_{ir}$  subfamily gene *KCNJ2* and the transcription factor Krüppel-like factor 5 (*KLF5*), both were found among the genes with EO-specific expression as well (Table 2). In the decreased expression group 3, there were two transmembrane ion transport genes (*KCNK6* and *KCNQ5*) and two cytoskeletal and sarcomeric genes (*ACTR3b* and *NHS*, Table 2).

**Table 2** Genes with expression correlated to EOD duration.

Group	Gene ID in annotation	Gene	Highlights of Predicted Function	Gene Description	Category
3	maker-ptg0003611-augustus-gene-0.2-mRNA-1	<i>ACTR3b</i>	F-actin dynamics / polymerization	ARP3 actin related protein 3 homolog B	cytoskeletal & sarcomeric
3	maker-ptg0005091-snap-gene-4.4-mRNA-1	<i>NHS</i>	regulator of actin remodelling	Nance-Horan syndrome protein	cytoskeletal & sarcomeric
3	maker-ptg0019661-augustus-gene-18.42-mRNA-1	<i>KCNK6</i>	outward rectification in a physiological potassium gradient and mild inward rectification in symmetrical potassium conditions	potassium channel subfamily K member 6	transmembrane ion transport
3	maker-ptg0006971-snap-gene-6.109-mRNA-1	<i>KCNQ5</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily Q member 5	transmembrane ion transport
5	maker-ptg0002651-est_gff_est2genome-gene-6.33-mRNA-1	<i>KCNJ2</i>	inwardly rectifying potassium channel	inward rectifier potassium channel 2	transmembrane transport
5	maker-ptg0017401-augustus-gene-1.112-mRNA-1	<i>KLF5</i>	transcription factor which might regulates potassium channel genes	krueppel-like factor 5	transcription factor

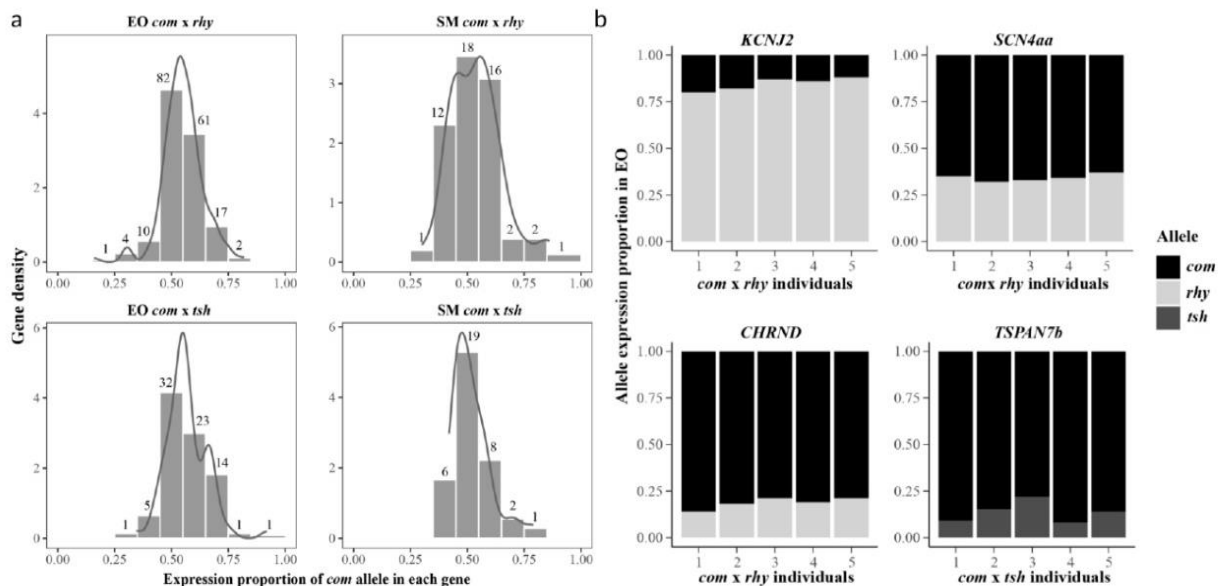
Assigning DEGs with increasing expression pattern to GO terms revealed 19 significantly enriched (Fisher's exact p-value < 0.05) GO terms (Supplementary Fig. 4, Supplementary Table 5). 18 genes were assigned to the enriched GO term "Golgi apparatus", 13 to "ion transport". Among the genes with decreasing expression pattern (group 3), 41 GO terms were significantly enriched (Supplementary Fig. 5, Supplementary Table 6). 12 of the genes were assigned to the GO term "axon guidance" which yielded the lowest p-value. There were also several enriched terms which might be functionally related to the EOD, e.g. membrane, Golgi membrane and apparatus, calcium ion binding, and ATP binding.

### Allele specific expression in F1 hybrids.

Two cohorts of F1 hybrids with one short duration EOD (*com* x *tsh*) and one medium duration EOD (*com* x *rhy*) were analyzed in our study. In total, we identified fixed SNPs (homozygous in parental species) in 177 genes differentially expressed in EO and in 77 differentially expressed in SM in the hybrid *com* x *rhy*. For the hybrid *com* x *tsh*, the respective SNP numbers were 52 in genes differentially expressed in the EO and 36 in genes differentially expressed in SM (Fig. 5a). For each of these genes, we calculated the allelic read proportion of the allele stemming from the parental species *com* (as identified by the fixed SNPs), averaged over the specimens of the respective hybrid cohort. In general, most genes exhibit an equal expression of both parental alleles,

with more genes have a *com* proportion near 0.5 (Fig. 5a). Among the genes with differentially expressed alleles, alleles stemming from *com* had an overall tendency towards higher expression, compared to the alleles from *rhy* or *tsh*, in both EO and SM from two hybrid cohorts (Fig. 5a).

In order to understand the allelic expression imbalance (AEI), we counted the number of genes with more than 0.6 proportion of one parental allele proportion for all individuals in each analyzed hybrid set<sup>31</sup>. In total, we identified 17 and 7 genes with AEI in EO and SM of the hybrid *com* x *rhy*, respectively; 2 and 1 such genes were identified in EO and SM of the hybrid *com* x *tsh*, respectively (Supplementary Table 7). In all the genes with AEI, the allele from *com* showed a higher expression proportion, except for the gene *KCNJ2* in the hybrid *com* x *rhy* where allele expression was biased in the opposite direction (average proportion of *com* allele was 0.16, Fig. 5b). We inferred amino acid sequences from the transcript sequences of the *KCNJ2* gene from *com*, *tsh* and *rhy*. The inferred protein sequence between *com* and *tsh* were identical, but *rhy* showed



**Fig. 5** Allele specific expression in electric organ (EO) and skeletal muscle (SM) among two hybrid cohorts *C. compressirostris* (*com*) x *C. rhynchophorus* (*rhy*) and *C. compressirostris* (*com*) x *C. tshokwe* (*tsh*).

**a** Gene density (y-axis) from two different tissues in each hybrids. The x-axis shows the expression proportion of the allele stemming from one parental species (*com*). Numbers above the bars represent the number of genes in the respective proportion ranges. **b** Proportion of two alleles from parental species in four genes related to ion transport and membrane (*KCNJ2*, *SCN4AA*, *CHRND* and *TSPAN7b*) in the EO. The x-axis represents the different individual samples of the corresponding hybrid cohort.

two amino acid substitutions at sites 60 (corresponding to the fixed SNP we identified in hybrids) and 198 (Supplementary Table 8). The amino acid substitution at site 60 was considered benign in the Polyphen2 analysis<sup>32</sup>, while the substitution at site 198 may have changed the protein function in *rhy* (inferred as probably damaging; Supplementary Table 9).

We also identified AEI in the gene *SCN4aa* in the EO of *com* x *rhy* hybrids (average proportion of *com* allele was 0.66, Fig. 5b). This proportion of the *com* allele is much higher than in the hybrid *com* x *tsh*, where the *com* allele of the *SCN4aa* only had a proportion of 0.46 in the EO. In the EO of hybrid *com* x *rhy*, AEI was identified in the gene *CHRND*, which might relate to ion channel gating (Fig. 5b). The *TSPAN6b* gene, encoding for an integral component of the plasma membrane, was also identified to exhibit a significant AEI in the EO of hybrid *com* x *tsh* (average proportion of *com* allele was 0.86, Fig. 5b).

## **Discussion**

### **Convergent gene expression in different electric fish lineages.**

The myogenic EO has convergently evolved six times in fishes. Even though the EOs show great differences in electrocyte morphology among independently evolved electric fish lineages, particular genes exhibit similar transcriptional expression patterns in the EO, relative to skeletal muscle<sup>9</sup>.

Several EO-specific candidate genes that we identified in *Campylomormyrus* were also overexpressed in the EO of other electric fish lineages, possibly indicating convergent expression pattern evolution in electric fish. This is particularly apparent in genes related to sodium and potassium currents. For instance, the  $Na_v$  channel gene *SCN4aa*, considered to be very important in regulating the sodium current to electrocytes, has been previously found overexpressed in the

EO of different electric fish lineages, i.e., Siluriformes, Gymnotiformes, and Mormyroidea other than *Campylomormyrus*<sup>6</sup>. The *FGF13a* that regulates this channel was consistently overexpressed in those electric fish species as well<sup>9</sup>. Interestingly, we identified another up-regulated ortholog (*FGF12*) in the EO which may have a similar function. In addition, multiple isoforms of sodium/potassium ATPase  $\alpha$  and  $\beta$  subunits and several transcription factors (*SIX2a*, *HEY1*) were found to be convergently up-regulated in the EO among these electric fish lineages<sup>9</sup>, a pattern confirmed for *Campylomormyrus* in our study.

Overexpression of another transcription factor (*MEF2a*) and of the calcium binding gene *SI00b* is characteristic for mormyrid EOs, i.e., *Paramormyrops*, *Brienomyrus*<sup>10,33</sup>, and *Campylomormyrus* (this study). We recently found *KCNA7a* to be tandemly duplicated in *Campylomormyrus*<sup>34</sup> and *Paramormyrops* (by re-analysis of the genome provided in<sup>33</sup>). This tandem duplication might be exclusive to mormyrid fishes, as we did not find it in available genomes neither of other electric fishes nor in *Scleropages* (a non-electric fish closely related to mormyrids; data not shown). In our study, both gene copies *KCNA7a\_1* and *KCNA7a\_2* were consistently up-regulated in the EO (*KCNA7a\_2* showed even higher expression than *KCNA7a\_1*). *KCNA7a* was inferred to be under positive selection in the transmembrane helices 3-4 linker, and is considered to relate to the differences in EOD duration among *Brienomyrus* and *Gymnarchus*<sup>23</sup>.

The *NDRG3* gene (N-Myc Downstream-Regulated Gene 3) exhibited a remarkable overexpression in the EO of *Campylomormyrus*. Interestingly, the phosphopeptides encoded by an ortholog (*NDRG4*) were highly enriched in the EO of the strongly discharging gymnotiform electric eel (*Electrophorus electricus*)<sup>35</sup>, indicating high expression level of this gene. In addition, *NDRG4* has been identified in zebrafish as a novel neuronal factor essential for sodium channel clustering at the nodes of Ranvier, the only places where action potentials are regenerated<sup>36</sup>. The function of



*NDRG3* in the nervous system has rarely been investigated. The *NDRG3* protein can interact with extracellular signal-regulated kinases (ERK1/2)<sup>37</sup>, which regulate *K<sub>v</sub>4.2* in the dendrites of hippocampal CA1 pyramidal neurons<sup>38,39</sup> as well as the *Nav1.7* channel<sup>40</sup>. In porcine as well as human lens, ERK1/2 is activated by the TRPV1 ion channel<sup>41</sup>, which was also overexpressed in EOs in *Campylomormyrus* (Supplementary Table 1).

### **Gene expression specificity in *Campylomormyrus*.**

The EO in mormyrids is derived from myogenic tissue, which transitions from a motoric/sarcomeric organization of muscle fibers to a continuous tube of electrocytes parallel to the spinal cord<sup>42</sup>. This transition process during the ontogeny of the EO involves cell size, morphology and physiology, and is still only partially understood. Some genes encoding for sarcomeric proteins, e.g. troponin I isoforms, myosin heavy chain and tropomyosin, are overexpressed in the EO of the mormyrid *Brienomyrus brachyistius*<sup>10</sup>, providing a preliminary insight into the developmental transition from SM to EO. In the EO of *Campylomormyrus*, however, we rarely found those genes up-regulated. Instead, the up-regulated actin-related genes in *Campylomormyrus* were more related to F-actin dynamics and included several unconventional myosins (Table 1, Supplementary Table 1). The four paralogous transcription factors *MEF2a-d* are responsible for the transcriptional activation of muscle-specific genes in the early specification of skeletal muscle<sup>43</sup>. Whereas *MEF2a* was overexpressed in the EOs of both *Brienomyrus*<sup>10</sup> and *Campylomormyrus*, a further paralog *MEF2b* was overexpressed only in *Campylomormyrus* (our study). The difference in expression of F-actin-related/sarcomeric proteins and *MEF2* transcription factors between two mormyrids genera suggests that the developmental transition in the EO might be different or, in other words, that the organization of the F-actin system in electrocytes may vary

across mormyrids. It has to be analyzed further whether these differences in the organization of the F-actin cytoskeleton concern the sarcomeric structure, the stalks of electrocytes, or both.

In addition, two paralogs of inwardly rectifying potassium channel ( $K_{ir}$ ) genes *KCNJ2* and *KCNJ9* were overexpressed in EOs of *Campylomormyrus*, along with *KCNQ5* and *KCNK2*. The mechanisms regulating potassium channels expression in electric fish are still unknown. We identified one transcription factor, Krüppel-like factor 5 (*KLF5*), that showed a high overexpression in EOs. In *Drosophila*, Krüppel is involved in the regulation of potassium channel expression. In case of a loss of *Shal* (*KCND*) potassium channel in *Drosophila*, Krüppel expression is induced and up-regulates expression of *Shaker* (*KCNA*) and *slowpoke* (*K<sub>ca</sub>*) potassium channels<sup>44</sup>. Remarkably, *Shal* (*KCND*) potassium channel is also not expressed in our studied *Campylomormyrus* species/hybrids<sup>44</sup>. We thus suppose that the EO differs from SM by the expression of a unique set of potassium channels that may contribute to the shape of the EO's action potential and thus the shape of the EOD signal. Moreover, we propose the *KLF5* gene to represent a transcription factor that drives the expression of regulating potassium channels in EO.

Our study has further revealed that different paralogs from the solute carrier family are active in EOs. Solute carriers form a group of membrane transport proteins located in various cellular membrane systems which transport diverse substrates including amino acids, oligopeptides, inorganic cations and anions<sup>45</sup>. We have found several genes of this gene family overexpressed in the EO that transport inorganic cations and anions, e.g. sodium, calcium, chloride. Especially the gene *SLC24a2* was highly overexpressed in *Campylomormyrus* EOs. This is a calcium/cation antiporter localized in the plasma membrane that mediates the extrusion of one calcium ion and one potassium ion in exchange for four sodium ions<sup>46</sup>. Overexpression of a calcium-extruding transporter in the EO indicates that regulation of the cytosolic calcium level in electrocyte differs

from that in SM. Unfortunately, we have no information yet on the distribution the *SLC24a2* protein within the electrocytes, and whether this calcium transporter is confined to a distinct region of the cell to mediate local regulation of the calcium level.

### **Differential gene expression with respect to EOD duration divergence among *Campylomormyrus* species.**

*Campylomormyrus* species produce species-specific EODs; their duration varies in a 100-fold range across species. The EOD is assumed to be mediated by sodium and potassium currents across the plasma membrane<sup>24</sup>. The depolarizing phase of an action potential is primarily produced by sodium influx. The repolarization phase is - along with a gradual decreasing sodium influx - affected by the orchestrated activities of delayed rectifier and inward rectifier potassium channels<sup>24,47</sup>. We suppose that species producing EODs of different duration may be equipped with different channel types or channel orthologs with different properties. However, certainly other mechanisms, such as different cell morphology, may also contribute to the EOD duration diversification.

The PCA from RNA-seq data showed a clear association between the overall gene expression and the EOD duration pattern (Fig. 3a). Based on the preliminary PCA and LRT result, we identified several genes which might contribute to EOD duration diversification in *Campylomormyrus*, including the potassium channel genes (*KCNJ2*, *KCNK6* and *KCNQ5*), actin-related genes (*ACTR3b*, *NHS*), and transcription factor *KLF5* (Table 2).

The gene *KCNK5* was found to be up-regulated in *Paramormyrops* (producing a short EOD) compared to the species with an elongated EOD<sup>48</sup>. In *Campylomormyrus*, the expression of another paralog *KCNK6* was also higher in species with short EOD. Two-pore potassium channels (*K<sub>2p</sub>*)

usually generate an outward potassium current and are also known as potassium leak channels. When silencing the *KCNK6* gene in the human heart, the action potential duration is prolonged<sup>49</sup>. Another voltage-gated potassium channel gene *KCNQ5* was decreasingly expressed in elongated EOD *Campylomormyrus* species. It forms M-type potassium current, a slowly activating and deactivating potassium conductance that works in determining the subthreshold electrical excitability of neurons<sup>50</sup>. The lower expression of both potassium channels genes in elongated EOD species will probably decrease the outward potassium current and consequently prolongate EOD repolarization.

The gene *KCNJ2* was increasingly expressed in elongated EOD species. It encodes for an inwardly rectifying potassium channel, with the greater tendency to allow potassium ions to flow into a cell rather than out of a cell<sup>50</sup>. The inward potassium current stabilizes the resting membrane potential of the cell and modulates the cardiac repolarization processes<sup>50,51</sup>. This inward rectifier channel-mediated potassium current is responsible for shaping the initial depolarization and final repolarization of the action potential in human cardiomyocytes<sup>52,53</sup>.

Regarding allele specific expression, there was a tendency towards higher expression of *com* alleles in the EOs and SMs of two analyzed hybrid cohorts (Fig. 5a). However, the phenotype of the EOD waveform in each hybrid is closer to the other parental species. This points towards some genes playing key roles in regulating the EOD waveform in the hybrids. The gene *KCNJ2* showed allelic expression imbalance (AEI) in *com* x *rhy* hybrids, which was the only gene with AEI and for which the *rhy* allele was preferentially expressed (Fig. 5b). The EOD in the adult hybrids *com* x *rhy* was of intermediate duration (4 ms), and the shape and waveform resemble the subadults' EOD in *rhy*. Both the EOD phenotype and the AEI in *KCNJ2* were hence closer to the parental species with the elongated EOD, i.e., *rhy*. The expression of *KCNJ2* in the EO among the purebred

species also increased with increasing EOD duration, e.g. the expression in *rhy* is higher than in *com*. This suggests that the *KCNJ2* gene might be under cis-regulation, and it should be a powerful candidate gene involved in the regulation of EOD duration in *Campylomormyrus*.

In addition, the *KCNJ2* gene in the species *rhy* (very long EOD) exhibits two non-synonymous substitutions, one of which predicted to cause a functionally relevant amino acid substitution (at site 198; Supplementary Table 8, 9). Interestingly, the same substitution at site 198 is present in another species with very long EOD (*C. numenius*, EOD duration 40 ms), while it is absent in other *Campylomormyrus* species with short or medium EOD which resemble the amino acid sequence of *com* and *tsh* (Cheri, Cheng & Tiedemann, unpubl. results). *C. numenius* and *rhy* are phylogenetically close<sup>54</sup>, such that the shared amino acid substitution could also reflect phylogenetic affinity. Nonetheless, the found amino acid substitution with inferred functional relevance could relate to the evolution of very long EODs in *Campylomormyrus*. Then, the *KCNJ2* gene could modulate EOD duration by a combination of expression level and functional protein sequence alteration. In summary, this study identifies the *KCNJ2*, *KCNK6* and *KCNQ5* genes, possibly in combination with other genes (e.g. *KLF5*, *ACTR3b*, *NHS*) as strong candidates underlying EOD duration diversification in the weakly electric fish genus *Campylomormyrus*. The diverged EOD likely affect the food spectrum and are used for mate recognition. This potential dual function in disruptive natural selection and pre-zygotic reproductive isolation would rank the EOD as a “magic trait”<sup>14</sup>, which may have promoted the ecological (probably sympatric) speciation and radiation of *Campylomormyrus* in the Congo River.

## Methods

**Animals, RNA isolation, library preparation and sequencing.** Three adult specimens of *C. tshokwe* were collected at Brazzaville/Republic of in the Congo River in 2012 and stored in

RNA later in  $-80^{\circ}\text{C}$ . Five adult specimens from each of the other two species (*C. compressirostris*, *C. rhynchophorus*) and two hybrids (*C. compressirostris* ♂ x *C. rhynchophorus* ♀, *C. compressirostris* ♂ x *C. tshokwe* ♀) were artificially bred and raised at the University of Potsdam. All specimens except for *C. tshokwe* were anesthetized by a lethal dose of clove oil, and dissected on cold 99% ethanol. Electric organ (EO) and skeletal muscle (SM) tissue from each specimen were flash frozen in liquid nitrogen, and further preserved in  $-80^{\circ}\text{C}$ . In total, we collected three samples of both EOs & SMs from *C. tshokwe*, five samples of both EOs & SMs from the other four species/hybrid cohorts in this study.

The RNA isolation was performed in all the EOs and SMs samples using QIAGEN RNeasy Fibrous Tissue Mini Kit. Total RNA concentration was estimated using a NanoDrop 1000 spectrophotometer (ThermoFischer Scientific, Germany), RNA quality was checked with an Agilent Bioanalyzer 2100 (Agilent Technologies, USA). mRNA enrichment was performed by poly (A) capture from isolated RNA using NEXTflex Poly (A) Beads. Strand-specific transcriptomic libraries were built using NEXTflex Rapid Directional RNA-Seq Kit (Bio Scientific, USA) based on the manufacturer's instructions.

Libraries were sequenced as 150 bp paired-end reads by Illumina HiSeq 4000 sequencing system at a commercial company (Novogene). Raw reads have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (accessions number:GSE240783). We trimmed the adapter sequences and low quality reads using a 4 bp sliding window with a mean quality threshold of 25, and a minimum read length of 36 bp by Trimmomatic v0.39<sup>55</sup>. Read quality, before and after read filtering, was measured by FastQC v0.11.9<sup>56</sup>.

**Differential gene expression analysis.** The quality-filtered reads from EOs and SMs were mapped to the *C. compressirostris* genome<sup>34</sup> using RSEM<sup>57</sup> for gene level-quantification estimation. The

estimated counts were imported into R/Bioconductor with the tximport package, which produced count matrices from gene-level quantification by taking the effective gene length into account<sup>58</sup>. Low count ( $\leq 10$ ) and low frequency (not present in at least two replicates) genes were removed. We performed a principle component analysis (PCA) from filtered and log-transformed counts. One SM sample from *C. compressirostris* was removed from this study, as its overall gene expression showed a deviant unusual pattern in the PCA.

We forwarded the normalized count matrices to DESeq2<sup>59</sup> to infer expression differences among EO and SM in each species/hybrid cohort respectively. We used a false discovery rate threshold of 0.05 to correct for multiple testing. The differentially expressed genes (DEGs) were identified with  $|\log_2 \text{ folder change } (\log_2\text{FC})| > 1$  and  $p\text{-value} < 0.05$ . In order to detect the EO specific gene expression pattern, we used Venn Diagrams<sup>60</sup> to visualize the shared DEGs (up- and down-regulated separately) among three purebred species and two hybrid cohorts.

The shared DEGs were annotated against the NCBI *nr* database by blastx with an e-value cutoff  $1e^{-10}$ . In addition, the up- and down-regulated DEGs in the EO were used to perform a Gene Ontology (GO) enrichment analysis<sup>27</sup>.

**RNA-seq data clustering by EOD duration.** The PCA plotting from log-transformed count matrices showed a clear pattern by the length of EOD (Fig. 3a). To identify genes with an expression pattern associated to EOD duration, we used DESeq2 to perform a likelihood ratio test<sup>61</sup> (LRT in the DESeq2 package). This test compares how well a gene's count data fit a "full model" compared to a "reduced model"<sup>29</sup>. Our full model was an equation:  $\text{full} = \sim \text{duration} * \text{tissue}$ . The duration is the length of the EOD in each purebred species, and tissue is the type of sample (EO or SM). The reduced model excluded the interaction between duration and tissue:  $\text{reduced} = \sim \text{duration} + \text{tissue}$ . Genes with adjust P-value ( $\text{padj}$ )  $< 0.05$  were considered to fit the "full model".

We used the `degPatterns` function from the ‘DEGreport’ package to cluster different groups with particular expression pattern using those significant genes across samples<sup>30</sup>, with `time = "duration"`, `col = "tissue"`.

The generated groups of different gene expression pattern across EOD duration were analyzed to identify genes with an expression pattern association with EOD duration. We hence focused on those groups fulfilling the following two criteria: 1) the gene expression level in EO is higher than SM in all F0 species; 2) the gene expression level in EO across EOD duration showed a consistent increasing or decreasing pattern.

The identified genes with increased and decreased expression relative to EOD duration were blasted against *nr* database using an e-value cutoff  $1e^{-10}$ . In addition, a GO term analysis was also performed for these genes.

**Allelic specific expression analysis.** The F1 hybrid contains two sets of subgenome from two parental species. Examination of allele specific expression can be applied to detect the allelic imbalance in transcription in heterozygous F1 hybrids. We only focused on transcripts of genes with biallelic SNPs fixed among the respective F0 parental species (hence, heterozygous only in F1 hybrids, homozygous in parental species).

We mapped the trimmed and filtered RNA-seq from five species/hybrids (in EOs and SMs, respectively) to the *C. compressirostris* genome using STAR v2.7.7<sup>62</sup>. The generated bam files were sorted according to the coordinates by SAMtools v1.15<sup>63</sup>. Variant calling was performed by BCFtools v1.9<sup>63</sup> in EOs and SMs respectively, using the command “`bcftools mpileup -f REFERENCE LIST_OF_BAM -Ou | bcftools call -mv -Ob -o BCFFILE`”, where the



REFERENCE, LIST\_OF\_BAM, BCFFILE were the CDS sequence name of the *C. compressirostris* genome, the list of bam files, and the output bcf file name, respectively.

After the variant calling, we performed the following steps to identify the fixed parental biallelic SNPs for each hybrid set. Firstly, we excluded the uncalled variants and only preserved biallelic SNPs using the command “bcftools view --exclude-uncalled -m2 -M2 BCFFILE > CALLING\_AD”, where the CALLING\_AD was the allelic depth for the final biallelic SNPs. Secondly, we discarded SNPs where the variant calling score at QUAL field was lower than 70, and allelic depth was lower than 10 in both alleles. Finally, we obtained high quality SNPs, at which both parental species were homozygous and fixed for a different allele.

For each hybrid, we calculated the expression proportion of the allele from *C. compressirostris* in EO and SM, respectively. We calculated the average proportion and its 95% confidence limits across biological replicates (and over SNPs in case of more than one SNP per locus; Fig. 5, Supplementary Table 7). Genes with *C. compressirostris* allele proportions <0.4 or >0.6 in the transcriptomes were considered exhibiting an imbalanced expression<sup>31</sup>.

### **Data availability**

Sequence data have been deposited at NCBI Gene Expression Omnibus under accession GSE240783.

### **References**

1. Darwin, C. *The origin of species*. (The Harvard Classics, 1859).
2. Bass, A. H. *Electric Organs Revisited: Evolution of a Vertebrate Communication and Orientation Organ. Electroreception*. (1986).
3. Kirschbaum, F. & Formicki, K. Structure and Function of Electric Organs. in *The Histology of Fishes* 75–87 (2020).

4. Smith, G. T. Evolution and hormonal regulation of sex differences in the electrocommunication behavior of ghost knifefishes (Apterontidae). *Journal of Experimental Biology* **216**, 2421–2433 (2013).
5. Glasauer, S. M. K. & Neuhauss, S. C. F. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genom.* **289**, 1045–1060 (2014).
6. Wang, Y. & Yang, L. Genomic Evidence for Convergent Molecular Adaptation in Electric Fishes. *Genome Biol Evol* **13**, 1–11 (2021).
7. Zakon, H. H. Adaptive evolution of voltage-gated sodium channels: The first 800 million years. *Proc Natl Acad Sci U S A* **109**, 10619–10625 (2012).
8. LaPotin, S. *et al.* Divergent cis-regulatory evolution underlies the convergent loss of sodium channel expression in electric fish. *Sci Adv* **8**, (2022).
9. Gallant, J. R. *et al.* Genomic basis for the convergent evolution of electric organs. *Science (1979)* **344**, 1522–1525 (2014).
10. Gallant, J. R., Hopkins, C. D. & Deitcher, D. L. Differential expression of genes and proteins between electric organ and skeletal muscle in the mormyrid electric fish *Brienomyrus brachyistius*. *Journal of Experimental Biology* **215**, 2479–2494 (2012).
11. Lamanna, F., Kirschbaum, F., Waurick, I., Dieterich, C. & Tiedemann, R. Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). *BMC Genom.* **16**, 1–17 (2015).
12. Kim, J. A., Jonsson, C. B., Calderone, T. & Unguez, G. A. Transcription of MyoD and myogenin in the non-contractile electrogenic cells of the weakly electric fish, *Sternopygus macrurus*. *Dev Genes Evol* **214**, 380–392 (2004).
13. Feulner, P. G. D., Plath, M., Engelmann, J., Kirschbaum, F. & Tiedemann, R. Electrifying love: Electric fish use species-specific discharge for mate recognition. *Biol Lett* **5**, 225–228 (2009).
14. Feulner, P. G. D., Plath, M., Engelmann, J., Kirschbaum, F. & Tiedemann, R. Magic trait electric organ discharge (EOD): Dual function of electric signals promotes speciation in African weakly electric fish. *Commun. Integr. Biol.* **2**, 329–331 (2009).
15. Tiedemann R., Feulner P.G.D., K. F. *Evolution in Action: Case studies in Adaptive Radiation, Speciation and the Origin of Biodiversity* 1–586 (2010) doi:10.1007/978-3-642-12425-9.
16. Paul, C. *et al.* 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 (2015).
17. Gallant, J. R., Arnegard, M. E., Sullivan, J. P., Carlson, B. A. & Hopkins, C. D. 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 (2011).

18. Korniienko, Y., Tiedemann, R., Vater, M. & Kirschbaum, F. Ontogeny of the electric organ discharge and of the papillae of the electrocytes in the weakly electric fish *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *Journal of Comparative Neurology* **529**, 1052–1065 (2021).

19. Kirschbaum, F. *et al.* Intra-genus (*Campylomormyrus*) and intergenus hybrids in mormyrid fish: Physiological and histological investigations of the electric organ ontogeny. *J Physiol Paris* **110**, 281–301 (2016).

20. Lamanna, F., Kirschbaum, F. & Tiedemann, R. 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 (2014).

21. Nagel, R., Kirschbaum, F. & Tiedemann, R. 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 (2017).

22. Paul, C., Kirschbaum, F., Mamonekene, V. & Tiedemann, R. Evidence for Non-neutral Evolution in a Sodium Channel Gene in African Weakly Electric Fish (*Campylomormyrus*, Mormyridae). *J Mol Evol* **83**, 61–77 (2016).

23. Swapna, I. *et al.* Electrostatic Tuning of a Potassium Channel in Electric Fish. *Curr. Biol* **28**, 2094–2102 (2018).

24. Stoddard P.K., M. M. R. Signal Cloaking by Electric Fish. *Bioscience* **58**, 415–425 (2008).

25. Mehaffey, W. H., Fernandez, F. R., Rashid, A. J., Dunn, R. J. & Turner, R. W. Distribution and function of potassium channels in the electrosensory lateral line lobe of weakly electric apteronotid fish. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **192**, 637–648 (2006).

26. Kuang, Q., Purhonen, P. & Hebert, H. Structure of potassium channels. *Cellular and Molecular Life Sciences* **72**, 3677–3693 (2015).

27. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, (2003).

28. Love, M. I., Anders, S., Kim, V. & Huber, W. RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Res* **4**, 1070 (2015).

29. Bendjilali, N. *et al.* Time-course analysis of gene expression during the *Saccharomyces cerevisiae* hypoxic response. *G3: Genes, Genomes, Genetics* **7**, 221–231 (2017).

30. L Pantano. DEGREport: Report of DEG analysis. *R package version* 1(8): 10.18129 (2019).

31. Chang, H. W. *et al.* Assessment of plasma DNA levels, allelic imbalance, and CA 125 as diagnostic tests for cancer. *J Natl Cancer Inst* **94**, 1697–1703 (2002).
32. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–249 (2010).
33. Gallant, J. R., Losilla, M., Tomlinson, C. & Warren, W. C. The genome and adult somatic transcriptome of the mormyrid electric fish *Paramormyrops kingsleyae*. *Genome Biol Evol* **9**, 3525–3530 (2017).
34. Cheng, F. *et al.* A new genome of an African weakly electric fish (*Campylomormyrus compressirostris*, Mormyridae) indicates rapid gene family evolution in Osteoglossomorpha. *BMC Genom.* **24**, 129 (2023).
35. Traeger, L. L., Sabat, G., Barrett-Wilt, G. A., Wells, G. B. & Sussman, M. R. A tail of two voltages: Proteomic comparison of the three electric organs of the electric eel. *Sci Adv* **3**, (2017).
36. Fontenas, L. *et al.* Neuronal Ndr4 Is Essential for Nodes of Ranvier Organization in Zebrafish. *PLoS Genet* **12**, 1–24 (2016).
37. Lee, D. C. *et al.* A lactate-induced response to hypoxia. *Cell* **161**, 595–609 (2015).
38. Schrader, L. A. *et al.* ERK/MAPK regulates the Kv4.2 potassium channel by direct phosphorylation of the pore-forming subunit. *Am J Physiol Cell Physiol* **290**, 852–861 (2006).
39. Gupte R P, Kadunganattil S, Shepherd A J, Merrill R, Planer W, Bruchas M R, Strack S, M. D. P. Convergent phosphomodulation of the major neuronal dendritic potassium channel Kv4. 2 by pituitary adenylate cyclase-activating polypeptide. *Neuropharmacology* **101**, 291–308 (2016).
40. Stambouliau, S. *et al.* ERK1/2 mitogen-activated protein kinase phosphorylates sodium channel Nav1.7 and alters its gating properties. *Journal of Neuroscience* **30**, 1637–1647 (2010).
41. Mandal A, Shahidullah M, D. N. A. TRPV1-dependent ERK1/2 activation in porcine lens epithelium Amritlal. *Exp Eye Res* **172**, 128–136 (2019).
42. Denizot, J. P., Kirschbaum, F., Westby, G. W. M. & Tsuji, S. 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 (1982).
43. Black, B. L. & Olson, E. N. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu Rev Cell Dev Biol* **14**, 167–196 (1998).
44. Parrish, J. Z. *et al.* Krüppel Mediates the Selective Rebalancing of Ion Channel Expression. *Neuron* **82**, 537–544 (2014).
45. He, L., Vasiliou, K. & Nebert, D. W. Analysis and update of the human solute carrier (SLC) gene superfamily. *Hum Genomics* **3**, 195–206 (2009).

46. Wang, L., Shao, Z., Chen, S., Shi, L. & Li, Z. A *SLC24A2* gene variant uncovered in pancreatic ductal Adenocarcinoma by whole exome sequencing. *Tohoku Journal of Experimental Medicine* **241**, 287–295 (2017).
47. Nass, R. D., Aiba, T., Tomaselli, G. F. & Akar, F. G. Mechanisms of Disease: Ion channel remodeling in the failing ventricle. *Nat Clin Pract Cardiovasc Med* **5**, 196–207 (2008).
48. Losilla, M., Luecke, D. M. & Gallant, J. R. The transcriptional correlates of divergent electric organ discharges in *Paramormyrops* electric fish. *BMC Evol Biol* **20**, 1–19 (2020).
49. Chai, S. *et al.* Contribution of two-pore K<sup>+</sup> channels to cardiac ventricular action potential revealed using human iPSC-derived cardiomyocytes. *Am J Physiol Heart Circ Physiol* **312**, H1144–H1153 (2017).
50. Hibino, H. *et al.* Inwardly rectifying potassium channels: Their structure, function, and physiological roles. *Physiol Rev* **90**, 291–366 (2010).
51. Li, M. *et al.* Overexpression of *KCNJ2* in induced pluripotent stem cell-derived cardiomyocytes for the assessment of QT-prolonging drugs. *J Pharmacol Sci* **134**, 75–85 (2017).
52. Dhamoon, A. S. & Jalife, J. The inward rectifier current (IK1) controls cardiac excitability and is involved in arrhythmogenesis. *Heart Rhythm* **2**, 316–324 (2005).
53. Jeevaratnam, K., Chadda, K. R., Huang, C. L. H. & Camm, A. J. Cardiac Potassium Channels: Physiological Insights for Targeted Therapy. *J Cardiovasc Pharmacol Ther* **23**, 119–129 (2018).
54. Lamanna, F. *et al.* Species delimitation and phylogenetic relationships in a genus of African weakly-electric fishes (Osteoglossiformes, Mormyridae, *Campylomormyrus*). *Mol Phylogenet Evol* **101**, 8–18 (2016).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
56. Andrew, S. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics* (2010).
57. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 1–16 (2011) doi:10.1201/b16589.
58. Paraskevopoulou, S., Dennis, A. B., Weithoff, G. & Tiedemann, R. Temperature-dependent life history and transcriptomic responses in heat-tolerant versus heat-sensitive *Brachionus* rotifers. *Sci Rep* **10**, 1–15 (2020).
59. Love, M. I., Anders, S. & Huber, W. *Differential analysis of count data - the DESeq2 package.* *Genome Biology* vol. 15 (2014).

60. Hanbo Chen, P. C. B. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* **12**, 1–7 (2011).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).
62. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* **51**, 11.14.1-11.14.19 (2015).
63. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).

### **Acknowledgements**

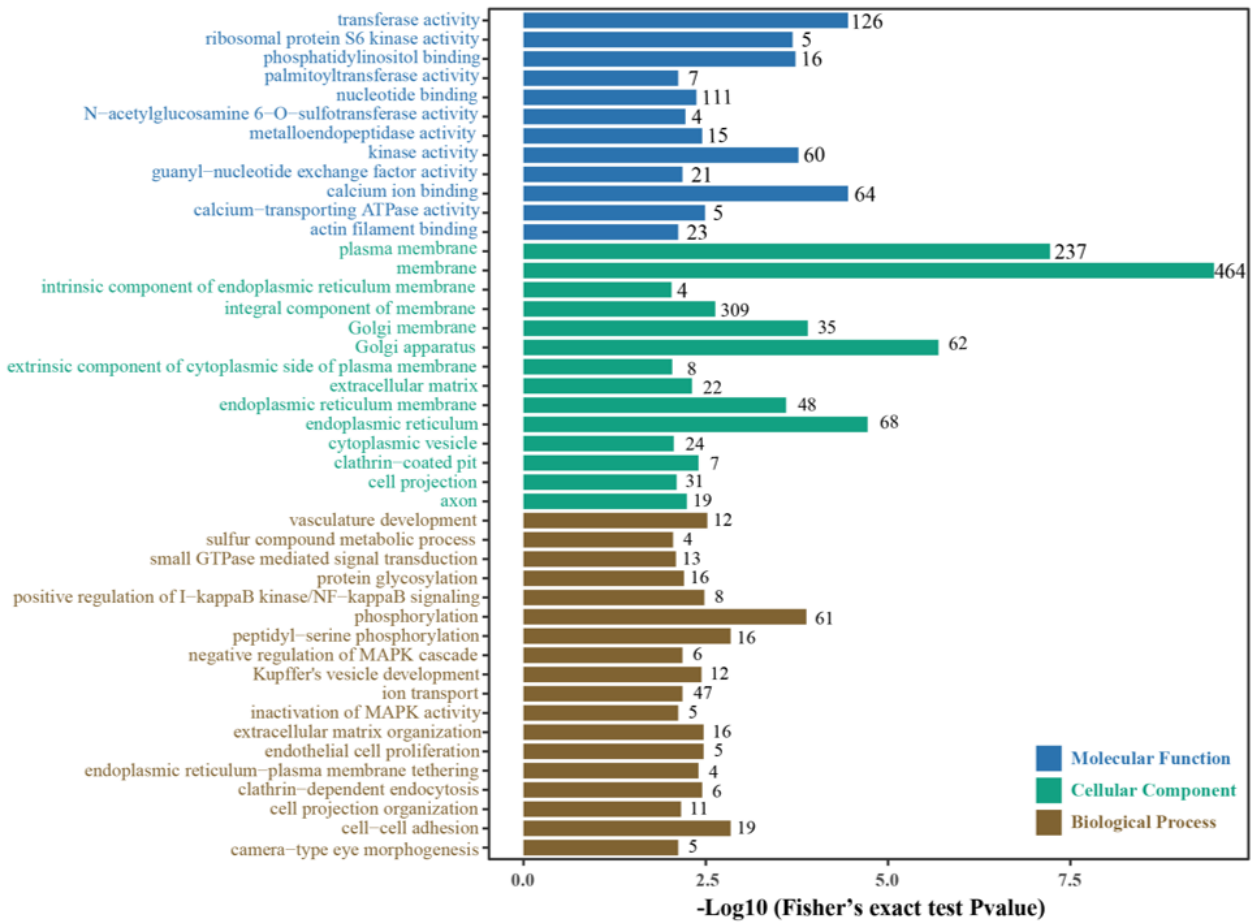
We thank Dr. Linh Nguyen for fish breeding and raising. This project is funded by University of Potsdam.

### **Competing interests**

The authors declare no competing interests.

## Supplementary information

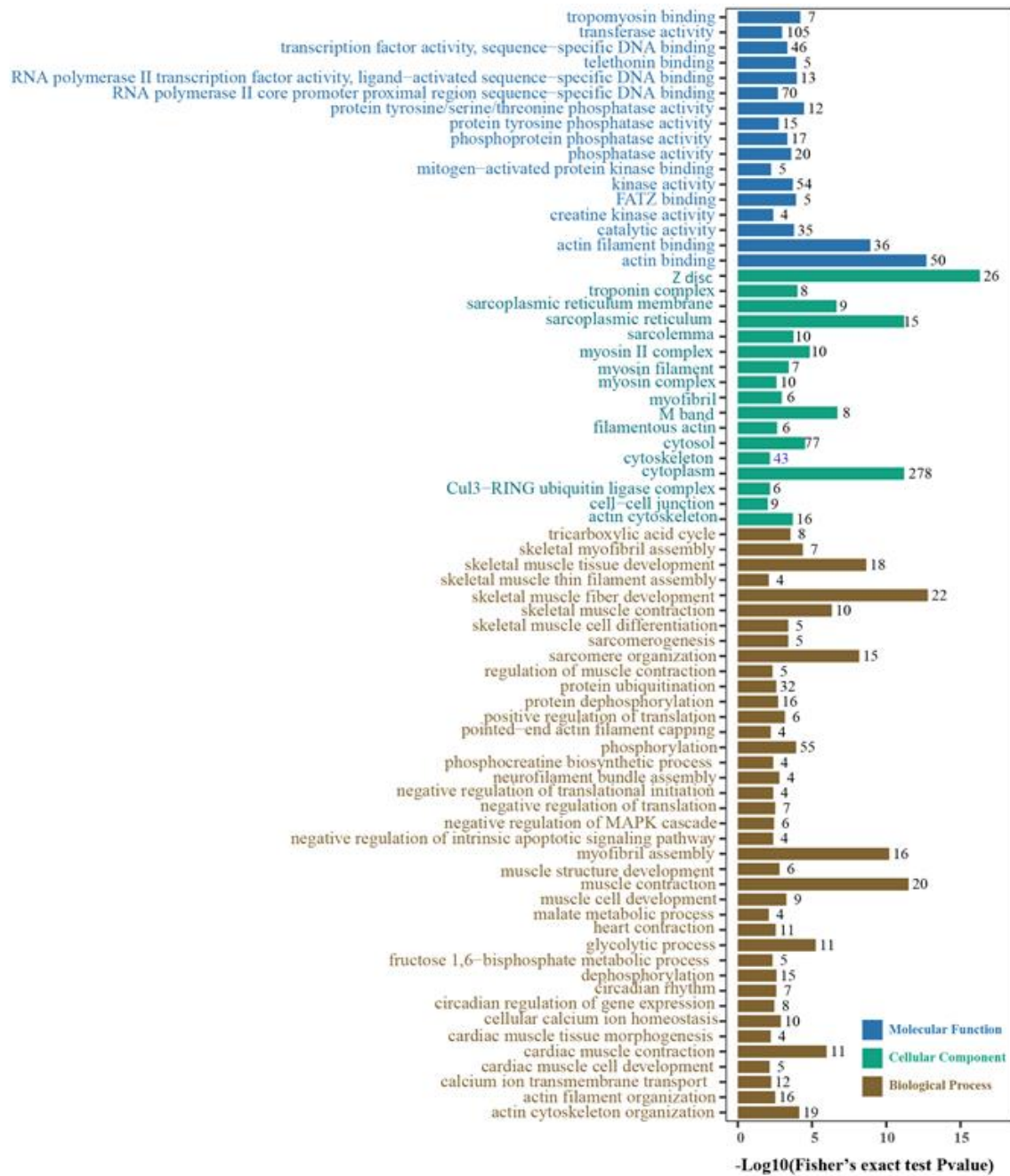
Supplementary Fig. 1



**Supplementary Fig. 1**

44 significantly enriched Gene Ontology (GO) terms with Fisher's exact test p-value < 0.01 in genes up-regulated in electric organ. The number of genes is plotted for each term. The GO terms are colored by their assignment to molecular function, cellular component, or biological process.

Supplementary Fig. 2

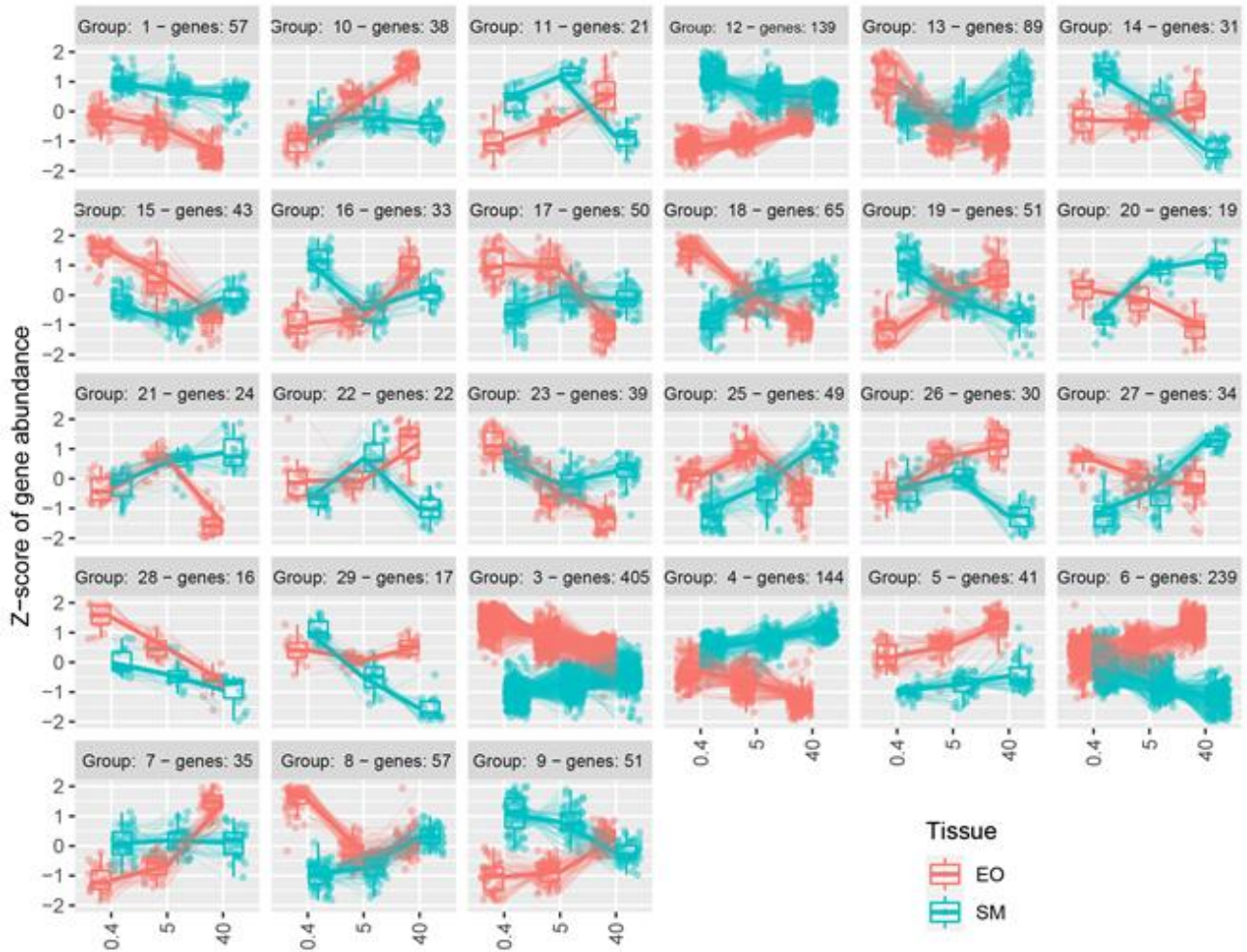


Supplementary Fig. 2

Significantly enriched Gene Ontology (GO) terms with Fisher's exact test P value < 0.01 of genes down-regulated in electric organ (up regulated in skeleton muscle). The number of genes is plotted for each term. The GO terms are colored by their assignment to molecular function, cellular component, or biological process.



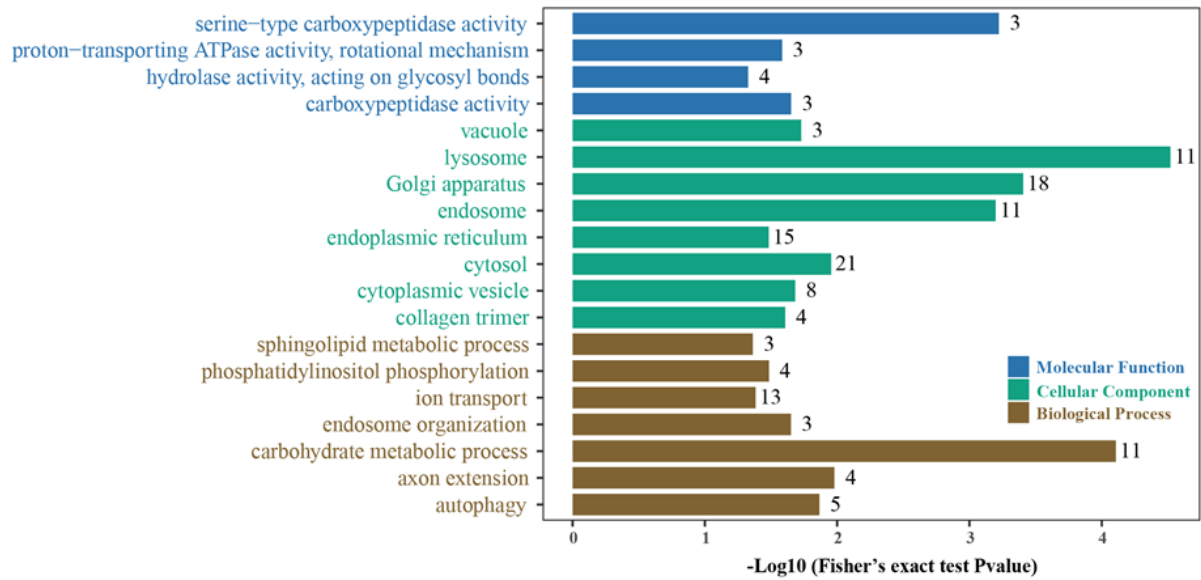
Supplementary Fig. 3



**Supplementary Fig. 3**

RNA-seq data clustering analysis based on EOD duration in electric organ (EO) and skeleton muscle (SM) of three F0 species. The x-axis for each group represents the EOD duration of the respective species: *C. compressirostris* (0.4ms), *C. tshokwe* (5ms) and *C. rhynchophorus* (40ms).

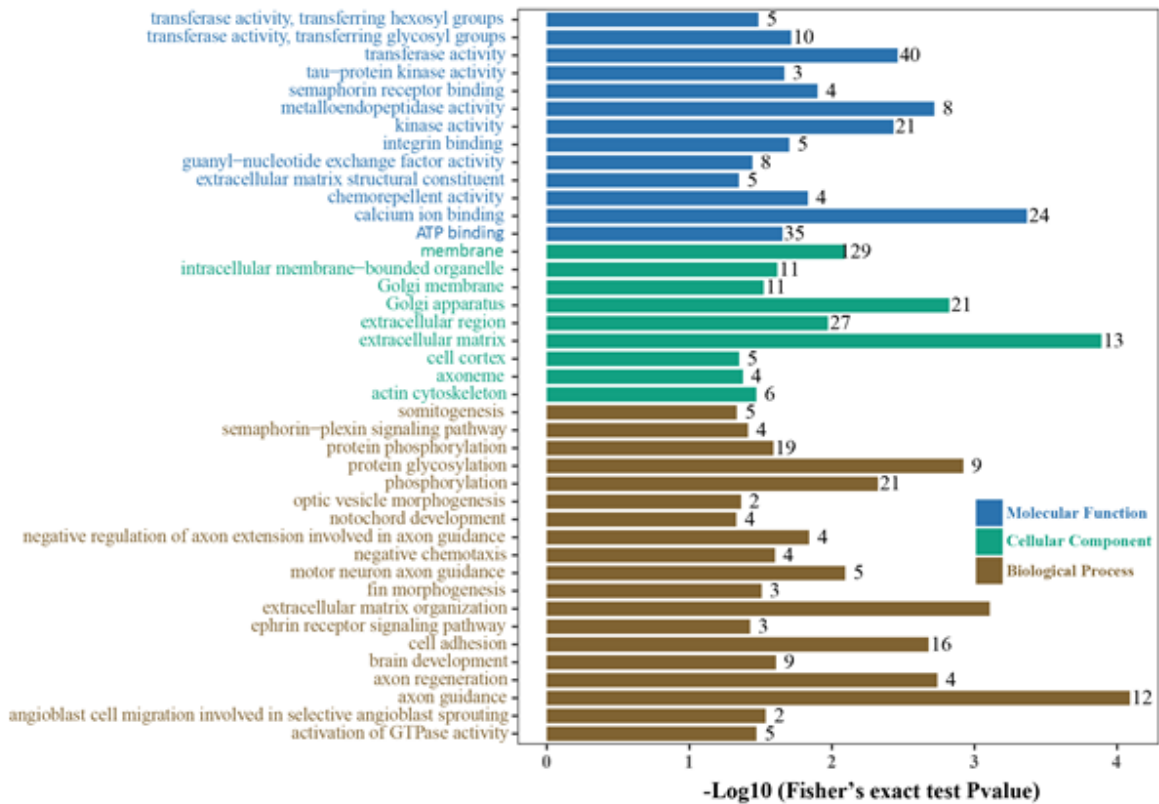
Supplementary Fig. 4



**Supplementary Fig. 4**

19 significantly enriched Gene Ontology (GO) terms with Fisher's exact test Pvalue < 0.05 for genes with increasing expression pattern over EOD duration change (Group 5 and 6). The number of genes is plotted for each term. The GO terms are colored by their assignment to molecular function, cellular component, or biological process.

Supplementary Fig. 5



**Supplementary Fig. 5**

Significantly enriched Gene Ontology (GO) terms with Fisher's exact test P value < 0.05 in genes with decreasing expression pattern over EOD duration change (Group 3). The number of genes is plotted for each term. The GO terms are colored by their assignment to molecular function, cellular component, or biological process.

## Supplementary Table 1

**Supplementary Table 1** Genes up-regulated in all species/hybrids in the electric organ relative to skeletal muscle.

ID	Blast Gene	Highlights of Predicted Function	Gene Description	Category	Average log2FC	Average Pvalue
maker-ptg0003611-augustus-gene-0.2-mRNA-1 snap_masked-ptg0009041-processed-gene-5.5-mRNA-1	<i>ACTR3b</i>	F-actin dynamics / polymerization	ARP3 actin related protein 3 homolog B	cytoskeletal & sarcomeric	3.02	3.1635E-20
maker-ptg0011801-augustus-gene-13.21-mRNA-1	<i>ARPC3</i>	F-actin dynamics / polymerization	actin related protein 2/3 complex subunit 3	cytoskeletal & sarcomeric	2.01	9.68947E-05
maker-ptg0020721-snap-gene-6.44-mRNA-1	<i>ARPC5</i>	F-actin dynamics / polymerization	actin-related protein 2/3 complex subunit 5	cytoskeletal & sarcomeric	1.99	0.003541395
maker-ptg0006501-snap-gene-7.49-mRNA-1	<i>CARMIL1</i>	F-actin dynamics / polymerization	capping protein regulator and myosin 1 linker 1	cytoskeletal & sarcomeric	1.91	0.001040325
maker-ptg0003461-snap-gene-25.173-mRNA-1	<i>EPPK1</i>	controls reorganization of intermediate filaments	epiplakin 1	cytoskeletal & sarcomeric	7.74	1.21807E-11
maker-ptg0012911-augustus-gene-1.16-mRNA-1	<i>GSN</i>	F-actin dynamics / polymerization	gelsolin	cytoskeletal & sarcomeric	6.70	2.7189E-50
maker-ptg0011821-snap-gene-5.74-mRNA-1	<i>MYO1d</i>	unconventional myosin; actin-based motor protein	unconventional myosin-Id	cytoskeletal & sarcomeric	2.16	0.013895309
maker-ptg0001901-snap-gene-4.78-mRNA-1	<i>MTSS1</i>	regulation of F-actin dynamics	metastasis suppressor protein 1	cytoskeletal & sarcomeric	4.71	1.09085E-10
maker-ptg0011881-snap-gene-5.24-mRNA-1	<i>MYH10</i>	unconventional myosin; actin-based motor protein	myosin-10	cytoskeletal & sarcomeric	1.57	2.93467E-10
maker-ptg0000281-processed-gene-137.57-mRNA-1	<i>MYLA</i>	regulatory light chain of myosin	myosin light chain 4	cytoskeletal & sarcomeric	7.12	5.64592E-05
maker-ptg0012911-snap-gene-1.34-mRNA-1	<i>MYO15a</i>	unconventional myosin; actin-based motor protein	unconventional myosin-XV	cytoskeletal & sarcomeric	3.68	3.40016E-09
maker-ptg0010031-snap-gene-2.16-mRNA-1	<i>MYO1d</i>	unconventional myosin; actin-based motor protein	myosin ID	cytoskeletal & sarcomeric	1.93	0.003762165
maker-ptg0020901-augustus-gene-35.16-mRNA-1	<i>MYO1e</i>	unconventional myosin; actin-based motor protein	unconventional myosin-Ie	cytoskeletal & sarcomeric	2.63	5.70736E-05
maker-ptg0005091-snap-gene-4.4-mRNA-1	<i>MYO3b</i>	unconventional myosin; actin-based motor protein	myosin-IIIb	cytoskeletal & sarcomeric	3.75	1.42712E-07
maker-ptg0021141-augustus-gene-0.11-mRNA-1	<i>NHS</i>	regulator of actin remodelling	Nance-Horan syndrome protein	cytoskeletal & sarcomeric	5.30	3.10465E-11
maker-ptg0006501-processed-gene-9.23-mRNA-1	<i>PARVG</i>	regulation of cytoskeleton organization	parvin gamma	cytoskeletal & sarcomeric	3.70	7.51203E-19
maker-ptg0000141-snap-gene-7.139-mRNA-1	<i>PLEC</i>	cross-linking and stabilization of cytoskeletal intermediate filaments	plectin	cytoskeletal & sarcomeric	2.62	8.77122E-22
maker-ptg0000561-processed-gene-18.17-mRNA-1	<i>MAP7d1</i>	microtubule cytoskeleton organization	MAP7 domain-containing protein 1	cytoskeletal & sarcomeric	8.32	6.90805E-34
	<i>ATP10d</i>	catalytic component of a P4-ATPase flippase complex	ATPase phospholipid transporting 10D	membrane organization	1.74	0.001111433

maker-ptg0002981-snap-gene-2.25-mRNA-1	<i>ESYT1</i>	ER-plasma membrane contact sites	extended synaptotagmin	membrane organization	2.30	3.56639E-09
maker-ptg0003171-snap-gene-0.11-mRNA-1	<i>CALR3b</i>	Ca2+-binding chaperone in ER	calreticulin	other	1.38	0.000700663
maker-ptg0003161-snap-gene-4.241-mRNA-1	<i>CAPN1a</i>	Ca2+-dependent, non-lysosomal cysteine protease	calpain-1 catalytic subunit	other	3.26	1.5809E-41
maker-ptg0002141-augustus-gene-8.178-mRNA-1	<i>CAPN5</i>	Ca2+-dependent, non-lysosomal cysteine protease	calpain 5	other	4.07	3.13077E-28
maker-ptg0008861-augustus-gene-0.134-mRNA-1	<i>CARHSP1</i>	regulation of mRNA stability	calcium regulated heat stable protein 1	other	3.01	2.85152E-08
maker-ptg0003161-snap-gene-7.47-mRNA-1	<i>CRIP3</i>	metal ion binding activity	cysteine rich protein 3	other	4.90	1.31827E-14
maker-ptg0006441-snap-gene-0.4-mRNA-1	<i>GDPD4</i>	enable metal ion binding activity and phosphoric diester hydrolase activity	glycerophosphodiester phosphodiesterase domain containing 4	other	4.98	3.97129E-12
maker-ptg0004051-snap-gene-38.13-mRNA-1	<i>SMOC2</i>	secreted calcium-binding protein	SPARC related modular calcium binding 2	other	5.40	1.01506E-17
maker-ptg0000701-snap-gene-5.18-mRNA-1	<i>FLOT2a</i>	scaffolding protein within caveolar membranes	flotillin-2a	other	5.77	1.81884E-53
maker-ptg0000401-snap-gene-14.53-mRNA-1	<i>NAT8l</i>	catalyzes the synthesis of N-acetylaspartate acid	N-acetyltransferase 8	other	7.97	1.96623E-07
maker-ptg0001021-snap-gene-52.7-mRNA-1	<i>ST8SIA5</i>	glycosphingolipid biosynthetic process	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5	other	9.67	6.85106E-08
maker-ptg0001201-snap-gene-2.54-mRNA-1	<i>SYNGR3</i>	synaptic vesicle membrane protein	synaptogyrin-3	other	10.00	3.33741E-07
maker-ptg0001521-snap-gene-4.203-mRNA-1	<i>ZDHHC23</i>	palmitoyltransferase	zinc finger DHHC-type containing 23	other	6.43	5.59154E-07
maker-ptg0018271-augustus-gene-0.22-mRNA-1	<i>FAT3</i>	cell-cell adhesion	FAT atypical cadherin 3	other	2.72	0.001238961
maker-ptg0001481-snap_masked-ptg0001481-processed-gene-1.18-mRNA-1	<i>FAT4</i>	cell adhesion molecule	FAT atypical cadherin 4	other	3.97	3.45897E-11
maker-ptg0014211-augustus-gene-5.0-mRNA-1	<i>JCAD</i>	cell adhesion	junctional protein associated with coronary artery disease	other	4.28	3.35791E-67
maker-ptg0014271-augustus-gene-38.97-mRNA-1	<i>KCTD9</i>	Substrate-specific adapter of a cullin-based E3 ubiquitin-protein ligase complex	potassium channel tetramerization domain containing 9	other	1.77	3.05442E-05
maker-ptg0003611-snap-gene-40.32-mRNA-1	<i>LAMA1</i>	extracellular matrix protein	laminin subunit alpha-1	other	5.73	3.75043E-57
maker-ptg0001871-augustus-gene-16.29-mRNA-1	<i>MMP28</i>	extracellular matrix protein	matrix metalloproteinase 28	other	4.82	6.22492E-59
maker-ptg0009641-snap-gene-1.24-mRNA-1	<i>NCAM1</i>	cell adhesion molecule	neuronal cell adhesion molecule	other	4.84	2.16487E-24
maker-ptg0007181-snap-gene-13.14-mRNA-1	<i>NDFIP2</i>	activates HECT domain-containing E3 ubiquitin-protein ligases	NEDD4 family-interacting protein 2	other	8.81	1.42211E-12
maker-ptg0008791-snap-gene-1.65-mRNA-1	<i>TMIGD1</i>	cell adhesion molecule	transmembrane and immunoglobulin domain containing 1	other	9.81	1.75298E-07
maker-ptg0012361-augustus-gene-24.16-mRNA-1	<i>TRH</i>	tripeptide hypothalamic regulatory hormone	thyrotropin releasing hormone	other	9.79	5.36476E-17

maker-ptg0007741-augustus-gene-4.234-mRNA-1	<i>CTNNAL1</i>	modulation the Rho pathway signaling	catenin alpha like 1	signaling	5.91	2.6796E-49
maker-ptg0000491-snap-gene-23.20-mRNA-1	<i>FGF12</i>	regulation of voltage-gated sodium channels	fibroblast growth factor 12	signaling	8.72	5.84977E-17
maker-ptg0000281-augustus-gene-28.23-mRNA-1	<i>HEG1</i>	calcium ion binding activity	protein HEG homolog 1	signaling	8.38	8.41625E-14
maker-ptg0001591-augustus-gene-6.0-mRNA-1	<i>PCP4</i>	modulator of calcium-binding by calmodulin	calmodulin regulator protein PCP4	signaling	6.64	4.01068E-05
maker-ptg0008691-snap-gene-20.89-mRNA-1	<i>PVALB9</i>	cytosolic Ca2+-binding protein of the EF-hand superfamily	parvalbumin, thymic	signaling	8.51	1.38958E-09
maker-ptg0002151-snap-gene-13.31-mRNA-1	<i>S100b</i>	cytosolic Ca2+-binding protein of the EF-hand superfamily	S100 calcium binding protein B	signaling	8.14	1.01224E-22
maker-ptg0012411-snap_masked-ptg0012411-processed-gene-1.62-mRNA-1	<i>KL</i>	may be involved in the regulation of calcium and phosphorus homeostasis	klotho	signaling	5.39	0.000196154
maker-ptg0008301-snap-gene-5.84-mRNA-1	<i>SEMA5a</i>	ligand for receptor PLXNB3	semaphorin-5A	signaling	9.17	2.9895E-15
maker-ptg0004081-augustus-gene-0.0-mRNA-1	<i>ANXA4</i>	annexin family of calcium-dependent phospholipid binding proteins	annexin A4	signaling	5.55	6.72209E-60
maker-ptg0000841-augustus-gene-28.11-mRNA-1	<i>CAMK1d</i>	Ca2+/calmodulin-dependent protein kinase	calcium/calmodulin dependent protein kinase ID	signaling	3.18	0.000145331
maker-ptg0002371-snap-gene-9.30-mRNA-1	<i>CAMK1g</i>	Ca2+/calmodulin-dependent protein kinase	calcium/calmodulin-dependent protein kinase type 1D	signaling	4.56	0.000111371
maker-ptg0000081-augustus-gene-0.0-mRNA-1	<i>CHRNB4</i>	neuronal acetylcholine receptor subunit alpha; nonselective cation channel	neuronal acetylcholine receptor subunit beta-4	signaling	6.28	2.08134E-05
maker-ptg0016101-snap-gene-4.137-mRNA-1	<i>GRIK3</i>	ionotropic glutamate receptor	glutamate ionotropic receptor kainate type subunit 3	signaling	6.15	5.72433E-05
maker-ptg0010291-snap-gene-17.22-mRNA-1	<i>GRIN2a</i>	ionotropic glutamate receptor	glutamate receptor ionotropic, NMDA 2A	signaling	6.25	0.001464578
maker-ptg0008971-snap-gene-2.61-mRNA-1	<i>GRINA</i>	negative regulation of apoptotic signaling pathway	glutamate ionotropic receptor NMDA type subunit associated protein 1	signaling	3.24	7.33209E-61
maker-ptg0000681-augustus-gene-18.3-mRNA-1	<i>ITPR1</i>	InsP3-dependent ER Ca2+ channel	inositol 1,4,5-trisphosphate receptor type 1	signaling	4.07	9.87654E-18
maker-ptg0008381-snap-gene-3.218-mRNA-1	<i>NDRG3</i>	predicted to be involved in signal transduction	N-myc downstream-regulated gene 3 protein	signaling	11.02	7.44456E-08
maker-ptg0014271-snap_masked-ptg0014271-processed-gene-2.9-mRNA-1	<i>P2RY2</i>	receptor for ATP and UTP coupled to G-proteins that activate a phosphatidylinositol-calcium second messenger system	P2Y purinoceptor 2	signaling	3.78	0.001473877
maker-ptg0002671-snap-gene-2.92-mRNA-1	<i>PIEZO2</i>	mechanosensitive ion channel	piezo type mechanosensitive ion channel component 2	signaling	1.89	1.03736E-06
maker-ptg0014241-augustus-gene-5.145-mRNA-1	<i>RET</i>	receptor tyrosine-protein kinase	ret proto-oncogene	signaling	10.14	2.32382E-12
maker-ptg0015631-augustus-gene-5.35-mRNA-1	<i>SGK1</i>	serine/threonine-protein kinase	serine/threonine-protein kinase Sgk1	signaling	7.79	6.63267E-38
maker-ptg0009261-snap-gene-6.19-mRNA-1	<i>TRPV1</i>	transient receptor potential family of ion channels; nociception	transient receptor potential cation channel subfamily V member 1	signaling	2.43	0.005919225

maker-ptg0010881-snap-gene-0.7-mRNA-1	<i>SIX2a</i>	target ARE promoter elements in Na <sup>+</sup> /K <sup>+</sup> adenosine triphosphatases	SIX homeobox 2	transcription factor	3.05	1.22103E-27
maker-ptg0007831-snap-gene-6.78-mRNA-1	<i>HEY1</i>	developing cardiac conduction pathway	hes related family bHLH transcription factor with YRPW motif 1	transcription factor	6.02	1.5325E-13
maker-ptg0001491-augustus-gene-2.87-mRNA-1	<i>ETV5</i>	transcription factor	ETS translocation variant 5	transcription factor	5.11	1.99403E-55
maker-ptg0007371-snap_masked-ptg0007371-processed-gene-6.55-mRNA-1	<i>FOXL2</i>	regulate distinct female sex determining pathways	forkhead box protein L2	transcription factor	8.52	4.14855E-07
maker-ptg0017401-augustus-gene-1.112-mRNA-1	<i>KLF5</i>	rebalance potassium channels	Krueppel-like factor 5	transcription factor	8.39	5.04913E-06
maker-ptg0000081-snap-gene-10.43-mRNA-1	<i>MEF2a</i>	transcriptional activator for numerous muscle-specific genes	myocyte-specific enhancer factor 2A	transcription factor	3.85	3.78175E-50
maker-ptg0012701-snap-gene-47.16-mRNA-1	<i>MEF2b</i>	transcriptional activator for numerous muscle-specific genes	myocyte-specific enhancer factor 2B	transcription factor	6.92	8.45833E-68
maker-ptg0003451-snap-gene-4.29-mRNA-1	<i>ANO10</i>	calcium-activated chloride channel	anoctamin 10	transmembrane ion transport	2.37	1.21792E-13
maker-ptg0004061-snap-gene-3.4-mRNA-1	<i>ANO5</i>	calcium-activated chloride channel	anoctamin 5	transmembrane ion transport	2.76	3.56495E-17
maker-ptg0009791-snap-gene-12.179-mRNA-1	<i>ANO6</i>	calcium-activated chloride channel calcium-activated nonselective cation (SCAN) channel which acts as a regulator of phospholipid scrambling	anoctamin 6	transmembrane ion transport	1.82	0.000445473
maker-ptg0009701-augustus-gene-2.127-mRNA-1	<i>ATP1a1</i>	Na/K-ATPase $\alpha$ -subunit	sodium/potassium-transporting ATPase subunit alpha-1	transmembrane ion transport	10.55	2.18943E-05
maker-ptg0011561-snap_masked-ptg0011561-processed-gene-0.19-mRNA-1	<i>ATP1a2a</i>	Na/K-ATPase $\alpha$ -subunit	sodium/potassium-transporting ATPase subunit alpha-2	transmembrane ion transport	3.07	2.63171E-12
maker-ptg0010471-snap-gene-1.63-mRNA-1	<i>ATP1b1a</i>	Na/K-ATPase $\beta$ -subunit	ATPase Na <sup>+</sup> /K <sup>+</sup> transporting beta 1a	transmembrane ion transport	6.59	3.63255E-59
maker-ptg0005091-snap-gene-9.39-mRNA-1	<i>ATP1b1b</i>	Na/K-ATPase $\beta$ -subunit	ATPase Na <sup>+</sup> /K <sup>+</sup> transporting beta 1b	transmembrane ion transport	4.51	2.06472E-06
maker-ptg0009931-snap-gene-3.19-mRNA-1	<i>ATP2a2</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 2	sarcoplasmic/endoplasmic reticulum calcium ATPase 2	transmembrane ion transport	1.90	5.79784E-08
maker-ptg0009041-snap_masked-ptg0009041-processed-gene-1.73-mRNA-1	<i>ATP2a2b</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 2	sarcoplasmic/endoplasmic reticulum calcium ATPase 2b	transmembrane ion transport	2.72	5.65072E-08
maker-ptg0000841-augustus-gene-36.7-mRNA-1	<i>ATP2b1a</i>	plasma membrane calcium-transporting ATPase 1	plasma membrane calcium-transporting ATPase 1a	transmembrane ion transport	2.96	4.08993E-15
maker-ptg0000841-snap-gene-36.24-mRNA-1	<i>ATP2b1b</i>	plasma membrane calcium-transporting ATPase 1	plasma membrane calcium-transporting ATPase 1b	transmembrane ion transport	3.66	1.07062E-27
maker-ptg0001351-augustus-gene-78.38-mRNA-1	<i>ATP2b3</i>	ATPase plasma membrane Ca <sup>2+</sup> transporting 3	ATPase plasma membrane Ca <sup>2+</sup> transporting 3	transmembrane ion transport	5.65	1.02377E-05
maker-ptg0003471-snap-gene-10.66-mRNA-1	<i>ATP2c1</i>	secretory pathway Ca <sup>2+</sup> -ATPase	ATPase secretory pathway Ca <sup>2+</sup> transporting 1	transmembrane ion transport	2.14	0.000146327
maker-ptg0000321-augustus-gene-1.26-mRNA-1	<i>ATP6ap2</i>	subunit of vacuolar-type H <sup>+</sup> -ATPase	ATPase H <sup>+</sup> transporting accessory protein 2	transmembrane ion transport	1.78	2.28333E-09
maker-ptg0005011-augustus-gene-3.39-mRNA-1	<i>ATP6v1f</i>	subunit of vacuolar-type H <sup>+</sup> -ATPase	ATPase H <sup>+</sup> transporting V1 subunit F	transmembrane ion transport	1.48	1.99856E-09

maker-ptg0014271-snap-gene-4.23-mRNA-1	<i>CACNA1b</i>	calcium channel subunit	voltage-dependent N-type calcium channel subunit alpha-1B	transmembrane ion transport	5.19	2.70467E-08
maker-ptg0011821-snap-gene-8.79-mRNA-1	<i>CNGB1</i>	nonselective cation channel	cyclic nucleotide-gated cation channel beta-1	transmembrane ion transport	5.76	2.75381E-05
snap_masked-ptg0000841-processed-gene-43.23-mRNA-1	<i>CRACR2aa</i>	Ca <sup>2+</sup> binding protein that is a key regulator of CRAC channel-mediated SOCE; adaptor protein for cytoplasmic dynein	calcium release activated channel regulator 2Aa	transmembrane ion transport	3.67	0.001092994
maker-ptg0010691-augustus-gene-3.65-mRNA-1	<i>GABRA1</i>	subunit of ligand-gated chloride channel	gamma-aminobutyric acid receptor subunit alpha-1	transmembrane ion transport	5.44	1.17436E-05
snap_masked-ptg0005091-processed-gene-7.11-mRNA-1	<i>GABRG3</i>	subunit of ligand-gated chloride channel	gamma-aminobutyric acid receptor subunit gamma-3	transmembrane ion transport	2.78	0.000289984
snap_masked-ptg0021791-processed-gene-1.70-mRNA-1	<i>GLRBB</i>	ligand-gated chloride channel	glycine receptor, beta b	transmembrane ion transport	9.03	2.58552E-29
maker-ptg0005591-snap-gene-2.76-mRNA-1	<i>HCN2</i>	hyperpolarization-activated ion channel	hyperpolarization activated cyclic nucleotide gated potassium and sodium channel 2	transmembrane ion transport	3.65	1.79576E-09
maker-ptg0000281-snap-gene-81.10-mRNA-1	<i>KCNA7a_1</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 7a	transmembrane ion transport	4.60	2.58472E-11
maker-ptg0000281-snap-gene-81.8-mRNA-1	<i>KCNA7a_2</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 7a	transmembrane ion transport	8.27	3.55527E-12
maker-ptg0014271-snap-gene-13.20-mRNA-1	<i>KCNIP3</i>	voltage-gated potassium channel	calsenilin	transmembrane ion transport	5.11	0.000993571
maker-ptg0002651-est_gff_est2genome-gene-6.33-mRNA-1	<i>KCNJ2</i>	inwardly rectifying potassium channel	inward rectifier potassium channel 2	transmembrane ion transport	5.53	1.12219E-20
maker-ptg0008301-augustus-gene-5.123-mRNA-1	<i>KCNJ9</i>	inwardly rectifying potassium channel	G protein-activated inward rectifier potassium channel 3	transmembrane ion transport	5.77	2.63245E-10
snap_masked-ptg0011181-processed-gene-0.13-mRNA-1	<i>KCNK2</i>	potassium two pore domain channel	potassium channel subfamily K member 2	transmembrane ion transport	6.15	9.23801E-14
maker-ptg0006971-snap-gene-6.109-mRNA-1	<i>KCNQ5</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily Q member 5	transmembrane ion transport	5.94	0.000938938
maker-ptg0011061-augustus-gene-0.48-mRNA-1	<i>MCOLN1</i>	intracellular cation channel	mucolipin-1	transmembrane ion transport	6.37	6.01855E-26
maker-ptg0010571-snap-gene-3.24-mRNA-1	<i>MCOLN3</i>	intracellular cation channel	mucolipin-3	transmembrane ion transport	5.68	4.30182E-06
maker-ptg0002531-augustus-gene-20.10-mRNA-1	<i>SCN1ba</i>	voltage-gated sodium channel beta-subunit (regulatory)	sodium channel subunit beta-1	transmembrane ion transport	3.84	1.29832E-07
maker-ptg0011881-snap-gene-6.4-mRNA-1	<i>SCN4aa</i>	voltage-gated sodium channel alpha-subunit	sodium channel protein type 4 subunit alpha A	transmembrane ion transport	10.98	1.27373E-11
maker-ptg0022391-snap-gene-5.5-mRNA-1	<i>SCN4b</i>	voltage-gated sodium channel	sodium voltage-gated channel beta subunit 4	transmembrane ion transport	5.34	8.18062E-47
maker-ptg0004771-augustus-gene-1.10-mRNA-1	<i>VDAC1</i>	ion channel in the outer mitochondrial membrane and also the outer cell membrane	voltage-dependent anion-selective channel protein 1	transmembrane ion transport	2.69	2.46626E-06
maker-ptg0000511-snap-gene-132.157-mRNA-1	<i>TMEM206</i>	proton-activated chloride channel	transmembrane protein 206	transmembrane ion transport	1.98	6.00534E-10
maker-ptg0013001-snap-gene-4.198-mRNA-1	<i>TMEM63c</i>	Ca <sup>2+</sup> permeable cation channel at ER/mitochondria contact sites	transmembrane protein 63C	transmembrane ion transport	4.00	0.001858763



maker-ptg0001871-augustus-gene-15.2-mRNA-1	<i>TMEM120a</i>	mechanosensing ion channel	transmembrane protein 120A	transmembrane ion transport	6.05	2.33605E-44
maker-ptg0001481-snap-gene-9.4-mRNA-1	<i>SLC24a2</i>	calcium, potassium:sodium antiporter	solute carrier family 24 member 2	transmembrane ion transport	9.06	1.1057E-37
maker-ptg0001021-snap_masked-ptg0001021-processed-gene-12.46-mRNA-1	<i>SLC4a4</i>	sodium bicarbonate cotransporter	solute carrier family 4 member 4	transmembrane ion transport	5.67	1.3753E-11
maker-ptg0000041-snap_masked-ptg0000041-processed-gene-3.9-mRNA-1	<i>SLC8a1a</i>	sodium/calcium exchanger probable lipid transporter that modulates cholesterol sequestration in the late endosome/lysosome	solute carrier family 8 member 1a	transmembrane ion transport	5.01	3.00748E-13
maker-ptg0022071-snap-gene-0.59-mRNA-1	<i>ABCA2</i>	ABC transporter; translocates drugs and phospholipids across the membrane	ATP binding cassette subfamily A member 2	transmembrane ion transport	2.21	1.89872E-10
maker-ptg0018441-snap-gene-4.2-mRNA-1	<i>ABCB1</i>	ABC transporter of broad substrate specificity	ATP binding cassette subfamily B member 1	transmembrane ion transport	4.72	0.000481822
maker-ptg0014361-snap-gene-0.52-mRNA-1	<i>ABCG2</i>	sodium:sulfate symporter	ATP-binding cassette sub-family G member 2	transmembrane ion transport	2.24	0.008846064
maker-ptg0003261-snap-gene-3.64-mRNA-1	<i>SLC13a1</i>	membrane transporter that exports free sialic acids	solute carrier family 13 member 1	transmembrane ion transport	3.41	3.99214E-05
maker-ptg0008781-snap_masked-ptg0008781-processed-gene-2.22-mRNA-1	<i>SLC17a5</i>	sodium/ascorbate cotransporter	solute carrier family 17 member 5	transmembrane ion transport	3.88	1.23265E-08
maker-ptg0014271-snap-gene-14.26-mRNA-1	<i>SLC23a2</i>	calcium-dependent mitochondrial solute carrier on the inner mitochondrial membrane	solute carrier family 23 member 2	transmembrane ion transport	2.45	3.35477E-05
maker-ptg0018371-snap_masked-ptg0018371-processed-gene-1.59-mRNA-1	<i>SLC25a23</i>	calcium-binding mitochondrial carrier on the inner mitochondrial membrane	solute carrier family 25 member 23	transmembrane ion transport	5.13	0.000247848
maker-ptg0001021-augustus-gene-64.1-mRNA-1	<i>SLC25a25b</i>	sodium ion- and chloride ion-dependent high-affinity transporter that mediates choline uptake	solute carrier family 25 member 25b	transmembrane ion transport	2.65	0.003639241
maker-ptg0001841-augustus-gene-2.1-mRNA-1	<i>SLC5a7</i>	sodium/glucose cotransporter 4	solute carrier family 5 member 7	transmembrane ion transport	5.91	0.000293221
maker-ptg0001751-augustus-gene-7.91-mRNA-1	<i>SLC5a9</i>	sodium-dependent vesicular transporter selective for proline, glycine, leucine and alanine	solute carrier family 5 member 9	transmembrane ion transport	2.80	6.19949E-11
maker-ptg0015871-snap-gene-9.16-mRNA-1	<i>SLC6a17</i>	sodium- and chloride-dependent transport of norepinephrine	solute carrier family 6 member 17	transmembrane ion transport	8.15	5.51716E-10
maker-ptg0008001-augustus-gene-2.11-mRNA-1	<i>SLC6a2</i>	sodium- and chloride-dependent transport of taurine and beta-alanine	solute carrier family 6 member 2	transmembrane ion transport	5.50	3.30148E-15
maker-ptg0010041-snap-gene-4.176-mRNA-1	<i>SLC6a6</i>		solute carrier family 6 member 6	transmembrane ion transport	4.10	7.32652E-09

## Supplementary Table 2

**Supplementary Table 2** Genes down-regulated in all species/hybrids in the electric organ relative to skeletal muscle.

<b>ID</b>	<b>Blast Gene</b>	<b>Highlights of Predicted Function</b>	<b>Gene Description</b>	<b>Category</b>	<b>Average log<sub>2</sub>FC</b>	<b>Average Pvalue</b>
snap_masked-ptg0003081-processed-gene-8.27-mRNA-1	<i>ABl2a</i>	regulator of actin cytoskeleton dynamics	abl interactor 2a	cytoskeletal & sarcomeric	-1.94	0.004180823
maker-ptg0005121-snap-gene-2.123-mRNA-1	<i>ABLIM1b</i>	binds to actin filaments and mediates interactions between actin and cytoplasmic targets	actin-binding LIM protein 1	cytoskeletal & sarcomeric	-2.88	1.69993E-09
maker-ptg0014581-snap-gene-2.2-mRNA-1	<i>ABRaa</i>	muscle specific actin-binding protein; involved in skeletal muscle hypertrophy and atrophy	actin-binding Rho-activating protein	cytoskeletal & sarcomeric	-5.52	1.11159E-06
maker-ptg0014161-augustus-gene-5.37-mRNA-1	<i>ABRab</i>	muscle specific actin-binding protein; involved in skeletal muscle hypertrophy and atrophy	actin binding Rho activating protein	cytoskeletal & sarcomeric	-5.37	5.05532E-07
maker-ptg0003061-augustus-gene-0.39-mRNA-1	<i>ACTA1</i>	actin isoform, striated muscle	actin, alpha 1, skeletal muscle	cytoskeletal & sarcomeric	-5.17	9.36728E-19
maker-ptg0010701-augustus-gene-5.71-mRNA-1	<i>ACTB1</i>	actin isoform, cytoplasmic	actin, cytoplasmic 1	cytoskeletal & sarcomeric	-1.95	0.00022477
maker-ptg0001711-snap-gene-9.31-mRNA-1	<i>ACTC1a</i>	actin isoform, striated muscle	actin, alpha cardiac muscle 1a	cytoskeletal & sarcomeric	-8.11	8.1603E-05
snap_masked-ptg0025191-processed-gene-0.5-mRNA-1	<i>ACTC1b</i>	actin isoform, striated muscle	actin, alpha cardiac muscle 1b	cytoskeletal & sarcomeric	-4.70	0.004466918
maker-ptg0000671-snap-gene-17.2-mRNA-1	<i>ACTC2</i>	actin isoform, striated muscle	actin, alpha, cardiac muscle 2	cytoskeletal & sarcomeric	-7.94	0.006969056
maker-ptg0005011-snap-gene-3.13-mRNA-1	<i>CAPZA2</i>	F-actin binding protein	F-actin-capping protein subunit alpha-2	cytoskeletal & sarcomeric	-2.35	1.10876E-11
maker-ptg0006931-snap-gene-21.128-mRNA-1	<i>CAPZB</i>	F-actin binding protein	F-actin-capping protein subunit beta	cytoskeletal & sarcomeric	-1.32	9.88642E-06
maker-ptg0001961-augustus-gene-4.0-mRNA-1	<i>KY</i>	cytoskeleton-associated protease required for normal muscle growth	kyphoscoliosis peptidase	cytoskeletal & sarcomeric	-7.80	0.000264434
maker-ptg0000071-snap-gene-2.65-mRNA-1	<i>MYBPC1</i>	thick filament-associated protein located in the crossbridge region of vertebrate striated muscle A bands	myosin binding protein C, slow type	cytoskeletal & sarcomeric	-6.60	0.000728724
snap_masked-ptg0000281-processed-gene-83.4-mRNA-1	<i>MYBPC2</i>	thick filament-associated protein located in the crossbridge region of vertebrate striated muscle A bands	myosin-binding protein C, fast-type	cytoskeletal & sarcomeric	-6.01	0.001321102
maker-ptg0003331-augustus-gene-6.128-mRNA-1	<i>MYBPC3</i>	thick filament-associated protein located in the crossbridge region of	myosin-binding protein C, cardiac-type	cytoskeletal & sarcomeric	-8.23	4.50551E-05

maker-ptg000962l-augustus-gene-12.1-mRNA-1	<i>MYBPH</i>	vertebrate striated muscle A bands binds to myosin; probably involved in interaction with thick myofilaments in the A-band	myosin-binding protein H	cytoskeletal & sarcomeric	-4.84	0.003814769
maker-ptg000085l-augustus-gene-17.7-mRNA-1	<i>MYH2</i>	unconventional myosin; actin-based motor protein	myosin heavy chain, fast skeletal muscle	cytoskeletal & sarcomeric	-8.17	5.93451E-19
snap_masked-ptg000407l-processed-gene-2.145-mRNA-1	<i>MYH1</i>	myosin heavy chain beta isoform expressed primarily in the heart, but also in skeletal muscles	myosin 1	cytoskeletal & sarcomeric	-8.60	4.34569E-06
maker-ptg000133l-augustus-gene-18.18-mRNA-1	<i>MYL1</i>	non-regulatory myosin light chain	myosin light chain 1	cytoskeletal & sarcomeric	-5.57	1.66255E-08
maker-ptg001175l-snap-gene-0.158-mRNA-1	<i>MYL2</i>	myosin regulatory light chain 2	myosin regulatory light chain 2, cardiac muscle	cytoskeletal & sarcomeric	-6.12	0.000536607
maker-ptg000187l-snap-gene-16.83-mRNA-1	<i>MYL2b</i>	myosin regulatory light chain 2	myosin regulatory light chain 2B, cardiac muscle	cytoskeletal & sarcomeric	-8.19	5.84804E-05
maker-ptg000100l-snap-gene-11.24-mRNA-1	<i>MYL4</i>	myosin regulatory light chain	myosin light chain 4	cytoskeletal & sarcomeric	-7.66	0.001195604
maker-ptg000135l-snap-gene-52.42-mRNA-1	<i>MYLK2</i>	myosin light chain kinase	myosin light chain kinase 2	cytoskeletal & sarcomeric	-4.77	0.000103916
maker-ptg000441l-augustus-gene-6.2-mRNA-1	<i>MYO16</i>	unconventional myosin; actin-based motor protein	unconventional myosin-XVI	cytoskeletal & sarcomeric	-4.85	0.006179819
maker-ptg000256l-snap-gene-3.25-mRNA-1	<i>MYO18a</i>	unconventional myosin; actin-based motor protein	unconventional myosin-XVIIIa	cytoskeletal & sarcomeric	-4.19	1.94298E-13
maker-ptg000960l-augustus-gene-0.14-mRNA-1	<i>MYO18b</i>	unconventional myosin; actin-based motor protein	myosin XVIIIb	cytoskeletal & sarcomeric	-3.94	8.08729E-15
maker-ptg000193l-snap-gene-2.47-mRNA-1	<i>MYOM1</i>	major component of the vertebrate myofibrillar M band	myomesin 1	cytoskeletal & sarcomeric	-7.91	1.13806E-07
maker-ptg000110l-augustus-gene-2.12-mRNA-1	<i>MYOM2</i>	major component of the vertebrate myofibrillar M band	myomesin 2	cytoskeletal & sarcomeric	-4.02	3.46432E-14
maker-ptg000405l-snap-gene-31.60-mRNA-1	<i>MYOZ1</i>	involved in linking Z-disk proteins	myozenin 1	cytoskeletal & sarcomeric	-6.71	6.15331E-07
maker-ptg000314l-augustus-gene-4.2-mRNA-1	<i>MYOZ2</i>	involved in linking Z-disk proteins	myozenin 2	cytoskeletal & sarcomeric	-7.19	0.001878985
maker-ptg000182l-snap-gene-3.82-mRNA-1	<i>MYOZ3</i>	involved in linking Z-disk proteins	myozenin 3	cytoskeletal & sarcomeric	-6.59	0.001769106
maker-ptg001112l-snap-gene-5.13-mRNA-1	<i>NEB</i>	binds and stabilize F-actin in the sarcomere	nebulin	cytoskeletal & sarcomeric	-7.40	0.000736895

maker-ptg000512l-snap-gene-3.14-mRNA-1	<i>NRAP</i>	may be involved in anchoring the terminal actin filaments in the myofibril to the membrane	nebulin related anchoring protein	cytoskeletal & sarcomeric	-3.92	1.29313E-19
maker-ptg002114l-snap-gene-0.37-mRNA-1	<i>PARVB</i>	adapter protein involved in the reorganization of the actin cytoskeleton	parvin beta	cytoskeletal & sarcomeric	-5.41	3.02009E-05
maker-ptg000572l-snap-gene-14.41-mRNA-1	<i>SMYHC1</i>	myosin heavy chain; actin-based motor protein	slow myosin heavy chain 1	cytoskeletal & sarcomeric	-7.57	0.004522565
maker-ptg000572l-augustus-gene-14.18-mRNA-1	<i>SMYHC2</i>	myosin heavy chain; actin-based motor protein	slow myosin heavy chain 2	cytoskeletal & sarcomeric	-5.20	0.001635808
maker-ptg000135l-augustus-gene-79.30-mRNA-1	<i>TNNC1</i>	component of troponin complex; regulation of muscle contraction	troponin C, slow skeletal and cardiac muscles	cytoskeletal & sarcomeric	-7.24	0.000307439
maker-ptg000135l-snap-gene-52.7-mRNA-1	<i>TNNC2</i>	component of troponin complex; regulation of muscle contraction	troponin C, skeletal muscle	cytoskeletal & sarcomeric	-4.81	5.01121E-11
maker-ptg000223l-snap-gene-2.48-mRNA-1	<i>TNNI2</i>	component of troponin complex; regulation of muscle contraction	troponin I, fast skeletal muscle	cytoskeletal & sarcomeric	-7.27	3.13868E-23
maker-ptg000445l-snap-gene-11.25-mRNA-1	<i>TPM1</i>	actin-binding protein; in association with the troponin complex involved in the striated muscle contraction	tropomyosin alpha-1 chain	cytoskeletal & sarcomeric	-4.81	6.55146E-15
maker-ptg001553l-augustus-gene-0.54-mRNA-1	<i>TPM2</i>	actin-binding protein; in association with the troponin complex involved in the striated muscle contraction	tropomyosin beta chain	cytoskeletal & sarcomeric	-6.67	1.53935E-06
maker-ptg001563l-snap-gene-12.53-mRNA-1	<i>TPM3</i>	actin-binding protein; in association with the troponin complex involved in the striated muscle contraction	tropomyosin alpha-3 chain	cytoskeletal & sarcomeric	-3.70	4.35071E-06
maker-ptg000555l-augustus-gene-11.55-mRNA-1	<i>TTN</i>	sarcomere organization	titin	cytoskeletal & sarcomeric	-2.95	3.70566E-07
snap_masked-ptg000176l-processed-gene-4.104-mRNA-1	<i>HSPB11</i>	disassembly of the sarcomeres	heat shock protein beta-11	cytoskeletal & sarcomeric	-4.89	5.99473E-07
maker-ptg001400l-snap-gene-3.115-mRNA-1	<i>CAPN3</i>	muscle-specific calpain; calcium-activated non-lysosomal thiol-protease	calpain 3	other	-4.51	3.46551E-17
maker-ptg000222l-snap-gene-9.11-mRNA-1	<i>CDH20</i>	calcium-dependent cell adhesion protein	cadherin-20	other	-3.77	0.000127412
maker-ptg000061l-augustus-gene-13.21-mRNA-1	<i>CDH26</i>	calcium-dependent cell adhesion protein	cadherin-like protein 26	other	-2.83	0.005372135
maker-ptg000534l-snap-gene-5.169-mRNA-1	<i>KCMF1</i>	E3 ubiquitin-protein ligase	potassium channel modulatory factor 1	other	-1.41	7.97951E-07
snap_masked-ptg002090l-processed-gene-16.20-mRNA-1	<i>ACOT12</i>	fatty acid metabolic process	acyl-CoA thioesterase 12	other	-7.79	0.00346148
maker-ptg001966l-	<i>CKM</i>	reversibly catalyzes the transfer of phosphate between	creatine kinase M-type	other	-8.43	5.01097E-09

augustus-gene-10.17-mRNA-1 maker- ptg0003931- augustus-gene-30.5-mRNA-1 maker- ptg0003231-snap- gene-1.83- mRNA-1 maker- ptg0002991-snap- gene-6.20- mRNA-1 snap_masked- ptg0011431- processed-gene-2.56-mRNA-1 maker- ptg0004871-snap- gene-5.15- mRNA-1 maker- ptg0002541- augustus-gene-6.0-mRNA-1 maker- ptg0006651-snap- gene-7.139- mRNA-1 maker- ptg0005581-snap- gene-4.101- mRNA-1 snap_masked- ptg0008691- processed-gene-20.110-mRNA-1 maker- ptg0009741- augustus-gene-10.59-mRNA-1 maker- ptg0028021-snap- gene-0.11- mRNA-1 maker- ptg0004811-snap- gene-2.8-mRNA-1 maker- ptg0001351- augustus-gene-82.84-mRNA-1 maker- ptg0010041- augustus-gene-9.10-mRNA-1 maker- ptg0002531- augustus-gene-58.60-mRNA-1 maker- ptg0000851-snap- gene-26.0- mRNA-1 maker- ptg0001711- augustus-gene-6.26-mRNA-1	<i>CKMT1a</i>	ATP and various phosphogens reversibly catalyzes the transfer of phosphate between ATP and various phosphogens	creatine kinase U-type, mitochondrial	other	-7.81	4.06172E-07
	<i>EPN2</i>	interacts with clathrin	epsin-2	other	-7.50	1.26567E-05
	<i>GLS2B</i>	glutaminase activity	glutaminase kidney isoform, mitochondrial	other	-7.18	1.23041E-05
	<i>IGFN1</i>	cell adhesion	immunoglobulin-like and fibronectin type III domain-containing protein 1	other	-7.80	5.98234E-36
	<i>LDHA</i>	catalyzes the conversion of L-lactate and NAD to pyruvate and NADH in the final step of anaerobic glycolysis	L-lactate dehydrogenase A chain	other	-7.47	0.000293053
	<i>TECR</i>	involved in both the production of very long-chain fatty acids	very-long-chain enoyl-CoA reductase	other	-7.82	0.000632574
	<i>MYOC</i>	secreted glycoprotein regulating the activation of different signaling pathways in adjacent cells	myocilin	signaling	-2.10	0.00200062
	<i>CALM3</i>	calcium-binding EF-hand protein	calmodulin 3	signaling	-2.84	8.29055E-05
	<i>PVALB2</i>	Ca <sup>2+</sup> -binding protein of the EF-hand superfamily	parvalbumin-2	signaling	-8.61	0.000478414
	<i>PVALB4</i>	Ca <sup>2+</sup> -binding protein of the EF-hand superfamily	parvalbumin 4	signaling	-6.32	0.000158436
	<i>PVALB7</i>	Ca <sup>2+</sup> -binding protein of the EF-hand superfamily	parvalbumin 7	signaling	-8.33	5.15658E-05
	<i>ADRB2</i>	beta-2-adrenergic receptor	adrenoceptor beta 2	signaling	-4.35	1.00955E-12
	<i>GPR173</i>	G-protein coupled receptor	probable G-protein coupled receptor 173	signaling	-8.04	0.00013504
	<i>SBK1</i>	serine/threonine kinase activity	serine/threonine-protein kinase SBK1	signaling	-6.93	0.00015683
	<i>MYOCD</i>	smooth muscle cells and cardiac muscle cells-specific transcriptional factor	myocardin	transcription factor	-3.89	1.03704E-11
	<i>MYOG</i>	muscle-specific transcription factor	myogenin	transcription factor	-2.73	1.67587E-05
	<i>ANO1</i>	calcium-activated chloride channel	anoctamin 1	transmembrane ion transport	-3.40	1.225E-17

maker-ptg000852l-snap-gene-16.100-mRNA-1	<i>ATP1b3a</i>	Na/K-ATPase $\beta$ -subunit	ATPase Na <sup>+</sup> /K <sup>+</sup> transporting subunit beta 3a	transmembrane ion transport	-4.10	1.54943E-19
maker-ptg000600l-augustus-gene-1.19-mRNA-1	<i>ATP2a1</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 1	sarcoplasmic/endoplasmic reticulum calcium ATPase 1	transmembrane ion transport	-7.30	3.1585E-07
maker-ptg000474l-snap-gene-6.56-mRNA-1	<i>ATP2a2a</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 2	sarcoplasmic/endoplasmic reticulum calcium ATPase 2a	transmembrane ion transport	-7.66	0.000724467
maker-ptg000085l-snap-gene-5.119-mRNA-1	<i>ATP2b2</i>	plasma membrane Ca <sup>2+</sup> transporting ATPase 2	ATPase plasma membrane Ca <sup>2+</sup> transporting 2	transmembrane ion transport	-2.46	0.002679005
maker-ptg000718l-snap-gene-53.80-mRNA-1	<i>ATP5f1b</i>	subunit of mitochondrial ATP synthase	ATP synthase subunit beta, mitochondrial	transmembrane ion transport	-1.81	2.19442E-06
maker-ptg000237l-snap-gene-5.34-mRNA-1	<i>CACNA1s</i>	voltage-gated calcium channel	dihydropyridine-sensitive L-type skeletal muscle calcium channel subunit alpha-1	transmembrane ion transport	-6.87	3.16021E-05
maker-ptg000085l-augustus-gene-14.68-mRNA-1	<i>CACNA2d1</i>	voltage-gated calcium channel	voltage-dependent calcium channel subunit alpha-2/delta-2	transmembrane ion transport	-4.46	1.86424E-11
maker-ptg000202l-snap-gene-1.0-mRNA-1	<i>CACNG1a</i>	voltage-gated calcium channel	voltage-dependent calcium channel gamma-1 subunit	transmembrane ion transport	-4.04	4.43468E-11
maker-ptg000763l-augustus-gene-3.165-mRNA-1	<i>CALHM6</i>	pore-forming subunit of a voltage-gated ion channel	calcium homeostasis modulator family member 6	transmembrane ion transport	-3.12	1.19758E-11
maker-ptg001156l-augustus-gene-8.0-mRNA-1	<i>CASQ1a</i>	calcium-binding protein in SR	calsequestrin-1a	transmembrane ion transport	-7.32	4.86051E-08
maker-ptg000021l-augustus-gene-6.30-mRNA-1	<i>CASQ1b</i>	calcium-binding protein in SR	calsequestrin-1b	transmembrane ion transport	-3.71	9.95212E-15
maker-ptg000650l-augustus-gene-22.51-mRNA-1	<i>CLCN1</i>	voltage-dependent chloride channel; important for repolarization of skeletal muscle cells after muscle contraction	chloride channel protein 1	transmembrane ion transport	-7.56	6.86007E-07
snap_masked-ptg000633l-processed-gene-25.11-mRNA-1	<i>KCNA1b</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 1b	transmembrane ion transport	-6.39	6.20958E-05
snap_masked-ptg000643l-processed-gene-7.150-mRNA-1	<i>KCNA4a</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 4a	transmembrane ion transport	-4.05	0.000266439
snap_masked-ptg000633l-processed-gene-25.9-mRNA-1	<i>KCNA5b</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 5b	transmembrane ion transport	-4.65	0.001218785
snap_masked-ptg000633l-processed-gene-26.11-mRNA-1	<i>KCNA6a</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 6a	transmembrane ion transport	-5.40	2.72077E-12
maker-ptg000600l-augustus-gene-13.28-mRNA-1	<i>KCNA7b</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily A member 7b	transmembrane ion transport	-7.39	0.000445521

maker-ptg0000721-snap-gene-0.6-mRNA-1	<i>KCNAB2</i>	voltage-gated potassium channel	voltage-gated potassium channel subunit beta-2	transmembrane ion transport	-3.37	9.48977E-05
maker-ptg0000681-snap-gene-8.7-mRNA-1	<i>KCNB1</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily B member 1	transmembrane ion transport	-5.58	0.006990731
maker-ptg0012481-est_gff_est2genome-gene-0.0-mRNA-1	<i>KCNE4</i>	voltage-gated potassium channel	potassium voltage-gated channel subfamily E regulatory subunit 4	transmembrane ion transport	-1.99	0.002653689
maker-ptg0003141-augustus-gene-2.9-mRNA-1	<i>KCNIP4</i>	voltage-gated potassium channel	Kv channel-interacting protein 4	transmembrane ion transport	-3.20	0.000884639
maker-ptg0013481-snap-gene-2.39-mRNA-1	<i>KCNK4</i>	potassium two pore domain channel	potassium channel subfamily K member 4	transmembrane ion transport	-4.94	0.00032563
maker-ptg0001701-snap-gene-11.9-mRNA-1	<i>KCNK7</i>	potassium two pore domain channel	potassium channel subfamily K member 1	transmembrane ion transport	-5.11	9.18843E-12
maker-ptg0002531-snap-gene-49.55-mRNA-1	<i>KCNN4</i>	calcium-activated potassium channel	potassium calcium-activated channel subfamily N member 4	transmembrane ion transport	-5.04	1.88707E-11
maker-ptg0007211-snap-gene-3.11-mRNA-1	<i>SCN3b</i>	voltage-gated sodium channel	sodium voltage-gated channel beta subunit 3	transmembrane ion transport	-2.11	5.45393E-10
maker-ptg0009741-augustus-gene-6.41-mRNA-1	<i>SCN4ab</i>	voltage-gated sodium channel	sodium channel protein type 4 subunit alpha b	transmembrane ion transport	-5.38	4.17695E-08
maker-ptg0004221-augustus-gene-0.23-mRNA-1	<i>SRL</i>	Ca <sup>2+</sup> -binding protein in SR; Ca <sup>2+</sup> buffering	sarcalumenin	transmembrane ion transport	-6.02	5.45506E-07
maker-ptg0005621-augustus-gene-8.257-mRNA-1	<i>TMEM38a</i>	monovalent cation channel in the SR and nuclear membranes of skeletal muscle	transmembrane protein 38A	transmembrane ion transport	-1.77	3.0994E-08
maker-ptg0006001-snap-gene-7.69-mRNA-1	<i>TRPM4</i>	Ca <sup>2+</sup> +-activated nonselective monovalent cation channel	transient receptor potential cation channel subfamily M member 4	transmembrane ion transport	-4.84	1.11494E-06
snap_masked-ptg0021141-processed-gene-2.64-mRNA-1	<i>ABCC9</i>	subunit of ATP-sensitive potassium channels	ATP-binding cassette sub-family C member 9	transmembrane ion transport	-3.10	3.68287E-09
maker-ptg0001811-snap-gene-1.66-mRNA-1	<i>SLC4a7</i>	sodium bicarbonate cotransporter	solute carrier family 4 member 7	transmembrane ion transport	-4.78	1.00112E-09
maker-ptg0003611-snap-gene-18.27-mRNA-1	<i>SLC22a23</i>	antiporter to transport organic ions across cell membranes	solute carrier family 22 member 23	transmembrane ion transport	-4.24	0.004015017
maker-ptg0007181-augustus-gene-55.59-mRNA-1	<i>SLC25a12</i>	mitochondrial electrogenic aspartate/glutamate antiporter	solute carrier family 25 member 12	transmembrane ion transport	-3.53	1.18714E-12
maker-ptg0002301-snap-gene-5.12-mRNA-1	<i>SLC41a1</i>	Na <sup>+</sup> /Mg <sup>2+</sup> ion exchanger	solute carrier family 41 member 1	transmembrane ion transport	-4.91	5.94485E-08

maker-ptg0006001-augustus-gene-10.17-mRNA-1	<i>SLC6a16</i>	Na(+)- and Cl(-)-dependent neurotransmitter transporter	solute carrier family 6 member 16	transmembrane ion transport	-2.21	1.60618E-05
maker-ptg0002991-snap-gene-9.30-mRNA-1	<i>SLC6a6b</i>	taurine:sodium symporter	solute carrier family 6 member 6b	transmembrane ion transport	-2.58	1.66873E-11

---



## Supplementary Table 3

Supplementary Table 3 44 Significantly enriched Gene Ontology terms with Fisher's exact test p-value < 0.01 in genes up-regulated in electric organ.

Term	GO terms	Category	Count	%	P-value	Genes	List Total	Pop Hits	Fold Enrichment	Bonfer mini	Bonfer roni	FDR mini	
GO:0016310	phosphorylation	Biological Process	61	5.41	1.32E-04	RET, SI:CH211-195B13.1, PFKFB2B, MOB1BA, PIK3CG, HK2, STK24B, RPS6KA3A, RPS6KA2, PIP5K1L, RPS6KA1, RPS6KB1B, AKT1, SI:CH211-243120.2, PIM3, ERBB4B, PDGFRA, PRKI, MAP4K3A, PRKCB, DCLK1A, SIDXEY-17214.3, EPHA4A, SHP, ACVR1BB, MAPK14A, PI4K2A, GNE, IPPK, ITPK1B, UCK2A, LTK, PRKX, TTK, CITA, PIP4K2AA, PRKC2, PFKPA, CKMT1, PIP5K1CA, GRK6, ERBB2, TRIOA, CAMK1GB, MAPK6, CHKB, YES1, TNK2B, LIMK2, PTK2AB, HIPK2, PRKACBB, ETNK2, JAK2B, RPS6KAL, PRKCHA, CMPK, PI4KB, PTK7B, CDK14, CAMK1DA	944	718	18397	1.66	0.20	0.22	0.22
GO:0018034	peptidyl-serine phosphorylation	Biological Process	16	1.42	0.001444	PRKI, SI:CH211-195B13.1, TTK, PRKX, TTBK1A, PRKC2, HIPK2, PRKCB, RPS6KA3A, RPS6KA2, RPS6KA1, RPS6KAL, RPS6KB1B, PRKCHA, CAMK1GB, CAMK1DA	944	122	18397	2.56	0.91	0.68	
GO:0030154	cell-cell adhesion	Biological Process	19	1.68	0.001444	DCHS1A, ARNT2, NEO1A, IGSF9BA, ITGA2B, CTNND1, NRCAMA, HSPG2, TMEM47, CNTN1A, DLG2, PERP, ITGA8, CDH13, ELMO2, ITGAV, PKP4, FAT4, CDH17	944	160	18397	2.31	0.91	0.68	
GO:0032759	vasculature development	Biological Process	12	1.06	0.003025	NOTCH2, YAP1, STN1, PANK2, ANGPT2B, CYP26C1, LGALS2A, AGAP2, ENPP2, ITGB8, RAB11A, MYCA	944	82	18397	2.85	0.99	0.68	
GO:0042882	positive regulation of kappab kinase/NF-kappab signaling	Biological Process	8	0.71	0.003316	PRKCB, CD40, TNIP2, TNFRSF19, RBCK1, S100B, LURAP1, MAP3K14A	944	39	18397	4.00	1.00	0.68	
GO:0042882	endothelial cell proliferation	Biological Process	5	0.44	0.00338	ITGA2B, ITGB8, ITGAV, ARHGEF7B, HSPG2	944	13	18397	7.50	1.00	0.68	
GO:0030154	extracellular matrix organization	Biological Process	16	1.42	0.003386	MMP15B, MMP17A, FBLN2, COLQ, SMOC2, COL4A2, ADAMTSL4, SI:DKY-6N6.1, COL1A1A, ADAMTSL3, ADAMTSL7, MMP28, COL2A1B, MMP19, ADAMTSL7, ADAMTSL9	944	133	18397	2.34	1.00	0.68	
GO:0070362	dathrin-dependent endocytosis	Biological Process	6	0.53	0.003582	AP2M1A, FCHO1, DNAJC6, SGIP1A, AP2A1, GPR107	944	21	18397	5.57	1.00	0.68	
GO:0070362	Kupffer's vesicle development	Biological Process	12	1.06	0.003658	YAP1, MBO3B, SNX10A, VGLL4B, ARL6, RAB3IP, DNMT3BB.1, ENPP2, ITGAV, GPR22A, RAB11A, ATP6V1F	944	84	18397	2.78	1.00	0.68	
GO:0060338	endoplasmic reticulum-plasma membrane tethering	Biological Process	4	0.35	0.004022	ESYT1A, ESYT2B, ESYT2A, GRAMD2AA	944	7	18397	11.14	1.00	0.68	
GO:0006468	protein glycosylation	Biological Process	16	1.42	0.006246	GALNT12, ST6GAL1, GALNT13, GALNT16, B3GAT2, EXT1B, FUT8A, ST6GALNAC5A, MGAT1B, B3GN17, ST3GAL4, ST8SIA5, ST8SIA6, STT3B, LARGE2, ST3GAL2	944	142	18397	2.20	1.00	0.80	
GO:0043009	negative regulation of MAPK cascade	Biological Process	6	0.53	0.006588	DUSP4, DUSP5, SPRED2B, PPEF2A, SPRED2A, DUSP7	944	24	18397	4.87	1.00	0.80	
GO:0006811	ion transport	Biological Process	47	4.17	0.006644	SIC24A2, GLRB, KCNGB3, SCN4BA, SCN1BA, TTYH2, PACC1, TMEM63C, ITPR1B, ATP2C1, ATP1A3A, CHRNG, ATP1B1B, ATP2A2B, KCNQB5, SIC39A7, GABRD, ATP1A1A.4, SI:CH211-225P5.8, ATP6V1F, SIC13A1, SCN4A4, HEPHL1B, GABRA1, CHRNB4, SIC8A1A, KCN19, SIC11A2, SIC39A10, TRPV1, SI:DKY-28B4.8, CNVGA3A, GRIN2AA, SIC5A9, ATP6V0A1A, KCN12A, KCN2A, SILCO4A1, ATP1A2A, ATP2B1A, ATP2B3B, CACNA1BA, VDACL1, SIC44A4, SIC44A8, MCU, GRIA3B	944	614	18397	1.49	1.00	0.80	
GO:0030154	cell projection	Biological Process	11	0.98	0.006953	CATIP, SNA10A, INTU, ARL6, CFL1, IFT122, TMEM237A, GPR22A, CCDC61, GSDA, SDCCAG8	944	79	18397	2.71	1.00	0.80	
GO:0006811	organization	Biological Process	5	0.44	0.00761	DUSP4, DUSP5, SPRED2B, SPRED2A, DUSP7	944	16	18397	6.09	1.00	0.80	
GO:0043009	activity	Biological Process	5	0.44	0.00761	SOX11A, LAMA1, LUM, ALDH1A2, IFT122	944	16	18397	6.09	1.00	0.80	
GO:0006811	camera-type eye morphogenesis	Biological Process	13	1.15	0.008039	RAC3A, BCAR3, ARHGAP32A, RHOB, DOCK4B, GDI1, DOCK7, TIAM1B, RHOF, RAPGEF11, RHOTB1, RHOCEB, RASGEF1BA	944	106	18397	2.39	1.00	0.80	
GO:0006811	small GTPase mediated signal transduction	Biological Process	4	0.35	0.008935	CHST7, CHST2B, SI:CH73-62B13.1	944	9	18397	8.66	1.00	0.84	
GO:0006811	sulfur compound metabolic process	Biological Process	4	0.35	0.008935	CHST7, CHST2B, SI:CH73-62B13.1	944	9	18397	8.66	1.00	0.84	

GO:001 membrane	464	41.1	3.42E-10	Cellular Component	983	7112	18868	1.25	0.00	0.00	0.00
6020											
GO:000 plasma membrane	237	21	6.08E-08	Cellular Component	983	3299	18868	1.38	0.00	0.00	0.00
5886											
GO:000 Golgi apparatus	62	5.5	2.02E-06	Cellular Component	983	628	18868	1.89	0.00	0.00	0.00
5794											
GO:000 endoplasmic reticulum	68	6.03	1.89E-05	Cellular Component	983	763	18868	1.71	0.01	0.00	0.00
5783											

GO:0000139	Cellular component	35	3.1	1.26E-04	GALNT12, SLC35B4, GALNT13, GALNT16, PGAP2, CLSTN1, PSENI2, RFNG, ATP2C1, ZDHHC4, KDELR2B, MAN2A2, CHST10, OSOX1, LARGE2, PCSK5B, CHST6, RHBDF1A, R1C1, CHST7, CAV3, CHST2B, B3GAT2, ZDHHC13, ZDHHC23B, ZDHHC17, ZDHHC14, SREBF2, ZDHHC8B, ZDHHC9, CDS1, ESYT1A, Si:DKEY-13N15.2, DIPK1B, NRROS, SLC35B4, PGAP2, EXT1B, PSENI2, FMN2B, ITPR1B, HACD3, ZDHHC4, KDELR2B, ABHD12, SEC61A1, CERS3A, ORMDL2, REEP3B, SC5D, SLC39A7, CALR3A, ESYT2A, SLC37A2, ICMT, CERS6, RHBDF1A, SDFZL1, ELOVL2, LMF2A, ATP6AP2, YIPF5, SCDB, ZDHHC14, SREBF2, RRBP1A, CYP51, BSCL2, G6PC3, ZDHHC9, CERS2B, DOLPP1, GRAMD18A, ERGIC3, SEC24D, PIGF, GRAMD18B	983	331	18868	2.03	0.05	0.01	0.01	0.01
GO:0005789	Cellular component	48	4.26	2.51E-04	CDS1, ESYT1A, Si:DKEY-13N15.2, DIPK1B, NRROS, SLC35B4, PGAP2, EXT1B, PSENI2, FMN2B, ITPR1B, HACD3, ZDHHC4, KDELR2B, ABHD12, SEC61A1, CERS3A, ORMDL2, REEP3B, SC5D, SLC39A7, CALR3A, ESYT2A, SLC37A2, ICMT, CERS6, RHBDF1A, SDFZL1, ELOVL2, LMF2A, ATP6AP2, YIPF5, SCDB, ZDHHC14, SREBF2, RRBP1A, CYP51, BSCL2, G6PC3, ZDHHC9, CERS2B, DOLPP1, GRAMD18A, ERGIC3, SEC24D, PIGF, GRAMD18B	983	529	18868	1.74	0.09	0.02	0.02	0.02
GO:0016021	Cellular component	309	27.4	0.00232	CRC3A2, SLC35B4, ZGC:165507, PGAP2, Si:DKEY-34D22.1, PIEKHB2, TMEEM200A, EXT1B, TXNDC11, ITPR1B, FRMD48A, ZDHHC4, NPDC1A, NSDHL, STS, SUTRK3A, ARL10, ESYT2B, ESYT2A, ENTPD1, ST6GAL1, CHST2B, SLC11A2, MIMP15B, ADGRA3, BCAM, SLC5A9, ATP1A2A, EPHA4A, GRAMD18A, MCOLN1A, GRAMD18B, FREM3, LRP1BB, CDS1, ESYT1A, RPN2, TMEEM63C, SGIPIA, ABCB5, MBOAT2B, PCDH19, HACD3, CRHR1, CD79B, CHST10, PDGFC, ANO5B, ST3GAL4, ST8SIA5, ST8SIA6, GSG1L2B, ANO10B, ST3GAL2, PRRT1, ABCC6A, ABCA2, CADM3, TMEEM86A, TMEEM184C, ICMT, VASNB, APCDD1L, Si:DKEY-15H18.17, NRXN3A, Si:CH211-286017.1, FLRT1B, EIF5, GPR180, NPTNB, MCOLN3A, TMEEM237A, NCAM1A, FAT4, PAM, AHTCF1, TENM4, TMEEM181, PHEX, CYB561D2, ZCCHC14, TMEEM47, ADAMTSL3, ATP2A2B, SMPD2B, ERBB4B, ATP1AJA.4, TMEEM198, Si:CH211-264F5.2, DCHS1A, CHST6, RHBDF1A, ZDHHC13, SLC39A10, TMEEM30AB, ANO6, MMEL1, GNLI, ZDHHC17, CSGALNACT1A, ZDHHC14, SREBF2, RGMA, FAM234B, B3GN7, SLCOM41, CRELD1, ERGIC3, GRINAA, DSCAM1L, PIGF, SLC5A7A, RPRMA, ZGC:92045, SLC16A6B, PACC1, HSD17B3, TMEEM72, GUCY2F, PPP1R3AA, TMEEM164, MEGF6A, ATP1A3A, GOPD4A, ORMDL2, PERP, MIEP1A.1, MUC13B, REEP3B, SLC17A5, KCNQ5B, SC5D, ZGC:162698, ZGC:92275, YIPF5, KLHL2, ZDHHC23B, PLXDC1, DPY19L3, ZDHHC8, CYB561, RCA2.1, RIC3B, TMEEM229B, GPR146, GALNT12, SEMA5A, TRHDE.1, KCNG3, GALNT13, NRROS, GALNT16, TUSC3, CLSTN1, ITGA2B, ACSL4A, COX6A1, TMEEM263, KDELR2B, ABHD12, LAPTM4B, TMEEM268, FAMI174B, RNFI9A, IGF1R1, Si:CH211-1E14.1, Si:DKEY-122A22.2, SLC39A7, ILL3RA2, LARGE2, ADAM19B, PDGFRA, SLC8A1A, SEMA6A, CCR12A, IGSF9BA, APLP2, SRD5A2A, SHISA3, ACPY1, SLC25A23B, FUT8A, FRMD3, TOM1L2, ALDH3A1, TRIM101, MAG, B3GLCTA, SYPL1, TMCC1B, FAM20C8, ABCG1, AQP1A.1, NOTCH2, DDX5, ABHD2A, DIPK1B, SCARB2A, TMEEM230A, PSENI2, Si:CH211-241B2.5, LDLRB, TMEEM241, TMEEM242, ADTRP1, MGAT1B, MAN2A2, CLDN11A, MFAP3L, KL, SLC37A2, ACSL3B, CRIM1, PLECA, GPR22A, SLC25A25B, CYP51, TMPRSS15, PTPRD, CNGA3A, IL17REL, G6PC3, BAXA, GPR107, SDK1A, STT3B, PTK7B, MUC1, PRR7, MTC1L, DYSF, NRCAMA, Si:CH73-2B9M14.2, MRC1A, HK2, BAGALNT3A, GRM4, Si:DKEY-12M2.1, Si:CH73-364H19.1, CERS3A, Si:DKEY-11F4.16, IL21R.1, CCDC51, GRAMD2AA, SLC13A1, HGSNAT, CERS6, ADGRV1, CHPF2, TNFRSF19, Si:CH211-153823.3, ATP6AP2, LMF2A, BRINP3A.1, NRG1, SLC25A55A, ST6GALNAC5A, SYNGR3A, ZDHHC8B, Si:DKEY-2884.8, TMIGD1, ATP6V0A1A, KCN2A, TMEEM218, EFHC1, NAT8L, CDH17, RNFL28A, CERS1, NEO1A, SCN48A, ATP10D, LRPS, MOSPDP2, PMP22B, ATP2C1, JAKMIP3, CLN3, CKMT1, CHSY1, ERBB2, TMEEM54A, CYB5A, LMBRD2B, KCN9, B3GAT2, BINIP3, ACKR4A, MBOAT1, MCOLN2, Si:DKEY-11F4.7, TMPPE, PNKD, FRAS1, CERS2B, MGAT4A, NDST1B, CNGB1A, AQP11, TRIM36, Si:CH211-152P11.8, SLC4A4A, SLC4A4B, PTPN5, CNM2B, CADM2B	983	5161	18868	1.15	0.59	0.13	0.13	0.12
GO:0005905	Cellular component	7	0.62	0.003965	AP2M1A, HIP1R8, FCHO1, SGIPIA, MYO6A, CLTCB, AP2A1	983	30	18868	4.48	0.78	0.19	0.18	
GO:0031012	Cellular component	22	1.95	0.004951	NRROS, VASNB, MIMP15B, MIMP17A, TSKU, COLQ, VCANB, CCN4A, COL4A2, ADAMTSL4, Si:DKEY-6N6.1, COL1A1A, ADAMTSL3, ADAMTSL7, Si:CH211-106H11.3, MIMP28, COL2A1B, TIMP2A, MIMP19, Si:DKEY-65B12.6, ADAMTSL7, ADAMTSL9	983	218	18868	1.94	0.85	0.21	0.20	
GO:0030424	Cellular component	19	1.68	0.005736	RET, SCN4AA, SYNM, SLC8A1A, SCN1BA, CLSTN1, ATP6AP2, ROGDI, ELAVL4, CCKA, NRCAMA, RAB11A, Si:DKEY-91M11.5, BCR, CNTN1A, DLG2, NPTNB, DSCAM1L, SLC5A7A	983	179	18868	2.04	0.89	0.22	0.21	
GO:0042995	Cellular component	31	2.75	0.007957	RAC3A, TENM4, INTU, ARL6, CLSTN1, ROGDI, ELAVL4, MYO6A, ABHD17C, RHOB1B, Si:DKEY-91M11.5, LAPTM4B, TEK13, RHOCB, ACTR3, PDGFRA, ADGRV1, ACTN1, IFT12, PTK2A8, RHOF, ZDHHC17, BCR, DLG2, TMEEM218, TMEEM237A, PI4K2A, GSNV, CCDC61, MACF1A, SDCCA68	983	361	18868	1.65	0.95	0.25	0.25	
GO:0031410	Cellular component	24	2.13	0.008806	RAC3A, Si:DKEY-13N15.2, DENND4C, ROGDI, SPRED2B, MYO6A, DYSF, ZDHHC13, RHOF, ZDHHC17, SREBF2, RHOB1B, KDELR2B, AP3M2, CADPSA, AP2M1A, TBC1D7, CLTCB, SPIRE1A, SEC24D, FLOT2A, RHOCB, CYB561, PI4K2A	983	259	18868	1.78	0.96	0.25	0.25	
GO:0031234	Cellular component	8	0.71	0.009172	ESYT1A, ESYT2B, ESYT2A, YES1, TNX2B, STAC3, PTK2AB, GRAMD2AA	983	46	18868	3.34	0.97	0.25	0.25	
GO:0031227	Cellular component	4	0.35	0.009321	ESYT1A, ESYT2B, ESYT2A, GRAMD2AA	983	9	18868	8.53	0.97	0.25	0.25	

GO:001 6740	transferase activity	126	11.2	3.52E-05	1743	17340	1.42	0.03	0.01	0.01	0.01
	Molecular Function				880	17340	1.42	0.03	0.01	0.01	0.01
					GALNT12, GALNT13, PRDM9, GALNT16, PFKFB2B, GTF2B, EXT1B, ZDHHC4, RPS6KA3A, RPS6KA2, RNF19A, RPS6KA1, AKT1, PIM3, LARGE2, PDGFRA, PRKCI, ST6GAL1, CHST2B, UBE2E3, MAP4K3A, FUT8A, RC3H2, DCLK1A, Si:CH211-259M1.8, B3GLCTA, EPHA4A, MAPK14A, RBCK1, PI4K2A, Si:CH73-62B13.1, GNE, CDS1, IPPK, UCK2A, ITPK1B, CRATA, UAP1, CITA, PIP4K2AA, PRKCZ, PKFPA, MGAT1B, PIP5K1CA, GRK6, CHST10, ST3GAL4, ST8SIA5, ST8SIA6, TRIOA, CAMK1GB, ST3GAL2, YES1, ICM1, CHK6, JAK2B, PRKCHA, NMT2, PI4KB, STT3B, PTK7B, CDK14, RET, PYGB, Si:CH211-195B13.1, MOB1BA, PIK3CG, HK2, B4GALNT3A, GYS1, STK24B, CPT2, CERS3A, PIP5K1L1, RPS6KB1B, Si:CH211-243J20.2, ERBB4B, CHST6, CHST7, CERS6, CHPT2, ELOVL2, ZDHHC17, ZDHHC14, CSGALNACT1A, ST6GALNAC5A, APRT, PRKCBB, ZDHHC8B, KAT2A, ERBB4B, CHST6, CHST7, CERS6, CHPT2, ELOVL2, ZDHHC17, ZDHHC14, CSGALNACT1A, ST6GALNAC5A, APRT, PRKCBB, ZDHHC8B, KAT2A, RET, SNED1, CAPN1A, CAPN1B, FKBP14, CETN2, PDCDC6, CLSTN1, CETN3, DYSF, ANXA11B, ITPR1B, EFHD2, ENPP2, EFHD1, EHD1B, DCHS1B, ESYT2B, ESYT2A, DCHS1A, EGF16, ACTN1, ANXA4, VWDE, SLC25A23B, HSPG2, MYL4, VCANB, RCN2, NID1A, CDHL3, CRELD1, PPEF2A, CRELD2, CDH17, GSNB, LRP1BB, FBLN7, ESYT1A, NOTCH2, EFS151JA, DIPK1B, NECA83, SWAP70B, LDLRB, PCDH19, FBLN2, MEGF6A, SPOCK3, REFS2, SLIT2, CALR3A, S100B, SLC25A25B, EDIL3A, MACF1B, SMO2C, PVALB8, FAT4, PCDHIG31, KCNIP3A, UNC138B, MACF1A, CACR2A						
GO:000 5509	calcium ion binding	64	5.67	3.59E-05	880	17340	1.71	0.03	0.01	0.01	0.01
	Molecular Function				RET, Si:CH211-195B13.1, PFKFB2B, MOB1BA, PIK3CG, HK2, STK24B, RPS6KA3A, RPS6KA2, PIP5K1L1, RPS6KA1, RPS6KB1B, AKT1, Si:CH211-243J20.2, PIM3, ERBB4B, PDGFRA, PRKCI, MAP4K3A, PRKCBB, DCLK1A, Si:DKY-172J4.3, EPHA4A, ACVR1BB, MAPK14A, PI4K2A, GNE, IPPK, ITPK1B, UCK2A, LTK, PRKX, TTK, CITA, PIP4K2AA, PRKCZ, PKFPA, CKMTI, PIP5K1CA, GRK6, ERBB2, TRIOA, CAMK1GB, MAPK6, CHK6, YES1, TNK2B, LIMK2, PTK2AB, HIPK2, PRKACBB, ETNK2, JAK2B, RPS6KAL, PRKCHA, CMPK, PI4KB, PTK7B, CDK14, CAMK1DA						
GO:001 6301	kinase activity	60	5.32	1.72E-04	880	17340	1.65	0.13	0.03	0.03	0.03
	Molecular Function				RET, Si:CH211-195B13.1, PFKFB2B, MOB1BA, PIK3CG, HK2, STK24B, RPS6KA3A, RPS6KA2, PIP5K1L1, RPS6KA1, RPS6KB1B, AKT1, Si:CH211-243J20.2, PIM3, ERBB4B, PDGFRA, PRKCI, MAP4K3A, PRKCBB, DCLK1A, Si:DKY-172J4.3, EPHA4A, ACVR1BB, MAPK14A, PI4K2A, GNE, IPPK, ITPK1B, UCK2A, LTK, PRKX, TTK, CITA, PIP4K2AA, PRKCZ, PKFPA, CKMTI, PIP5K1CA, GRK6, ERBB2, TRIOA, CAMK1GB, MAPK6, CHK6, YES1, TNK2B, LIMK2, PTK2AB, HIPK2, PRKACBB, ETNK2, JAK2B, RPS6KAL, PRKCHA, CMPK, PI4KB, PTK7B, CDK14, CAMK1DA						
GO:003 5091	phosphatidylinositol binding	16	1.42	1.87E-04	880	102	17340	3.09	0.14	0.03	0.03
	Molecular Function				ESYT1A, ESYT2B, ESYT2A, ARHGAP32A, Si:CH211-195B13.1, ITPR1B, PITPNBL, SNX21, STAM, ZCCHC14, TOM1L2, SH3YL1, SNX10A, HIP1RB, SNX7						
GO:000 4711	ribosomal protein S6 kinase activity	5	0.44	2.03E-04	880	7	17340	14.07	0.15	0.03	0.03
	Molecular Function				RPS6KA3A, RPS6KA2, RPS6KA1, RPS6KAL, RPS6KB1B						
GO:000 5388	calcium-transporting ATPase activity	5	0.44	0.003246	880	13	17340	7.58	0.93	0.40	0.40
	Molecular Function				Si:DKY-2884.8, ATP2B1A, ATP2A2B, ATP2B3B, ATP2C1						
GO:000 4222	metallopeptidase activity	15	1.33	0.003522	880	122	17340	2.42	0.94	0.40	0.40
	Molecular Function				MMP15B, MIMEL1, PAPPAA, MMP17A, PHEX, PITRM1, ADAMTSL4, ADAMTSL3, ADAMTSL7, MEJJA.1, MMP2B, MMP19, ADAMTSL7, ADAMTSL9, ADAM19B						
GO:000 0166	nucleotide binding	111	9.84	0.004302	880	1703	17340	1.28	0.97	0.43	0.43
	Molecular Function				ADCV1A, PANK2, TAOK2A, NUBP2, RPS6KA3A, RPS6KA2, RPS6KA1, DHX58, AKT1, PIM3, EHD1B, PDGFRA, PRKCI, ENTDPD1, UBE2E3, MAP4K3A, AARS1, DCLK1A, Si:CH211-256M1.8, PRKAR1B, EPHA4A, ATP1A2A, ATP2B1A, MAPK14A, ROR2, PI4K2A, ABCG1, IPPK, UCK2A, ITPK1B, DDX5, RTEL1, DHX8, ARL6, RRAD, ABCB5, MYO6A, TUBA8L3, TUBA8L2, CITA, PIP4K2AA, PRKCZ, PKFPA, PIP5K1CA, GRK6, TRIOA, CAMK1GB, MYH10, RHOCB, Si:DKY-32E23.4, ABCG6A, ABCA2, RAB4B, YES1, MYO15A, EIF5, JAK2B, PRKCHA, PI4KB, CDK14, RALAA, KIF13BA, RET, Si:CH211-195B13.1, HSP90A1, MCM7, HK2, KIF15, ARL5C, PIP5K1L1, ATP2A2B, RPS6KB1B, Si:CH211-243J20.2, ERBB4B, ATP1A1A.4, MBD3B, GNLI, PRKCBB, Si:DKY-2884.8, UBE2R2, Si:DKY-172J4.3, ACVR1BB, EEF1A1B, BLVRA, LTK, ATP10D, PRKX, TTK, ATP2C1, ADCY7, CKMT1, ATP1A3A, RASD1, Si:CH211-257P13.3, ERBB2, ARF3A, MAPK6, KIF26BA, MAP4K4, TNK2B, MYO1EA, PTK2AB, HIPK2, PRKACBB, ATP2B3B, RPS6KAL, ABCG2A, CMPK, KRAS, CAMK1DA, RAN						
GO:000 1517	N-acetylglucosamine 6-O-sulfotransferase activity	4	0.35	0.006	880	8	17340	9.85	0.99	0.51	0.51
	Molecular Function				CHST6, CHST7, CHST2B, Si:CH73-62B13.1						
GO:000 5085	guanyl-nucleotide exchange factor activity	21	1.86	0.006592	880	215	17340	1.92	0.99	0.51	0.51
	Molecular Function				BCAR3, ARHGFE10, RIC1, DOCK4B, IQSEC3A, DENND4C, RAB31P, DOCK7, TIAM1B, RAB31L1, ARHGFE7A, ARHGFE7B, ARHGFE7C, Si:DKY-91M11.5, NET1, BCR, MAD2, TRIOA, SBF1, RASGEF1BA						
GO:005 1015	actin filament binding	23	2.04	0.007522	880	247	17340	1.83	1.00	0.51	0.51
	Molecular Function				ACTR3, MBD3B, MYO1EA, FHOD1, ACTN1, MYO6A, MYO15A, NEB, SHROOM1, ARPCS1A, CORO2BA, DUB, PSTPIP1A, SAMD14, HIP1RB, ARPC3, CFL2, CFL1, TWNF1B, GAS2L1, CTNNA1, MYH10, GSNB						
GO:001 6409	palmitoyltransferase activity	7	0.62	0.007671	880	35	17340	3.94	1.00	0.51	0.51
	Molecular Function				ZDHHC8B, ZDHC9, ZDHC13, ZDHC23B, ZDHC17, ZDHC14, ZDHC4						

## Supplementary Table 4

Supplementary Table 4 76 Significantly enriched Gene Ontology terms with Fisher's exact test p-value < 0.01 in genes down-regulated in electric organ.

Term	GO terms	Category	Count	%	P-value	Genes	List	Pop	Pop	Fold	Bonferroni	Benjamini	FDR	
GO:004	skeletal muscle	Biological	22	2.26	1.52E-13	CAVIN4B, KLHL41A, MYBP1C, PYGMA, RBFOX2, SMPX, MYO18AB, RBFOX1L, LGALS2A, MYO18AA, SIX1B, RYR3, ACTN2B, LMOD3, RYR1B, RYR1A, NFIXA, KLHL40B, KLHL40A, KLHL41B, SMYD1B, MYF5	820	63	18397	7.835	2.09E-10	2.09E-10	2.06E-10	
8741	fiber development	Process												
GO:000	muscle contraction	Biological	20	2.06	3.14E-12	TMOD1, TNN1C, MYH8, TNNI4A, TMOD4, TPM3, TPM1, TNN1B, TNN1A, TNN1C, MYOM1A, MYOM1B, LMOD3, LMOD2B, SPEGB, TNNI2A.4, TNN1Z, TPMA, DESMA, TNNI2A.1	820	58	18397	7.736	4.33E-09	2.16E-09	2.13E-09	
6936	GO:003	myofibril assembly	Biological	16	1.65	6.69E-11	TMOD1, TMOD4, TNN1B, LMOD3, LMOD2B, TTN.2, TNNI2A.4, TTN.1, MEZF2A, MEZF2B, TTN.2	820	40	18397	8.974	9.22E-08	3.07E-08	3.02E-08
0239	GO:000	skeletal muscle	Biological	18	1.85	2.25E-09	PROX1A, CRYABA, DESMA, PGM5, SMYD1B	820	65	18397	6.213	3.10E-06	7.75E-07	7.61E-07
7519	tissue	Process				MYOG, POPDC3, CDKN1A, DNAJB6, MYLPP8, FHL1A, STAC3, FXR1, SYNPO2LA, SYNPO2LB, TTN.2, BAG3, TTN.1, NFIXA, CRYABA, DESMA, ITGA7, MYF5								
GO:004	sarcomere	Biological	15	1.54	6.83E-09	KLHL41A, CAPN3A, LRRC39, TFP12, TNN1B, TNN1A, ACTN2B, SIMYHC2, TTN.2, TNN1Z, MYH7L, DESMA, FLNCB, KLHL41B, SMYD1B	820	46	18397	7.316	9.42E-06	1.88E-06	1.85E-06	
5214	organization	Process												
GO:000	skeletal muscle	Biological	10	1.03	4.70E-07	TNN1C, RYR1B, TNNI4A, ZMP:0000000930, TNNI2A.4, TNNC1B, RYR1A, STAC3, TNNI2A.1, TCAP	820	24	18397	9.348	6.48E-04	1.08E-04	1.06E-04	
3009	contraction	Process												
GO:006	cardiac muscle	Biological	11	1.13	1.05E-06	SIMYHC2, TNN1C, MYL13, TNNI4A, ZMP:0000000930, TNNI2A.4, TNNC1B, MYH7L, TNNI2A.1, TNN1A, TCAP	820	33	18397	7.478	0.001452	2.08E-04	2.04E-04	
0048	contraction	Process												
GO:000	glycolytic process	Biological	11	1.13	5.68E-06	PFKMB, GPIB, INSRA, PGAM2, PKMA, TPI1B, ALDOA, ENO3, ALDOAB, GAPDH, ALDOCB	820	39	18397	6.328	0.007795	9.78E-04	9.61E-04	
6096	GO:001	skeletal myofibril	Biological	7	0.72	4.17E-05	MYO18AB, TPM3, MYO18AA, TMOD4, TTN.1, DUSP27, SMYD1B	820	16	18397	9.815	0.055863	0.00638703	0.00627
4866	assembly	Process												
GO:003	actin cytoskeleton	Biological	19	1.95	7.32E-05	PDLIM5B, PDLIM3B, PHACTR3B, ACTN3B, EHBPL11A, EHBPL11B, ACTN2B, SSH2A, ROCK2A, CAPZB, STARD13B, DAAM2, CAPZA1B, LDB3B, XIRP1, FINA, CORO1CA, ZGC:162952, SMTNL1	820	144	18397	2.960	0.096053	0.01009811	0.00992	
0036	organization	Process												
GO:001	phosphorylation	Biological	55	5.66	1.18E-04	COQ8AA, MYLK2, DYRK4, CDKN1A, PRKAB1A, MAST2, CKMT2A, AKAP8L, CAMK2N1A, ROCK2A, EEF2K, ULK1B, MYLK4A, PLAU, PIK3R3B, ADKB, GRK7A, PRKG1B, ADKA, ZGC:172076, HUNK, BMP1A, MAPKAPK3, PIK3CA, PRKCO, MET, UCKL1B, VEGFAA, DAPK2A, RAF1A, AK1, CITA, PAK1, INSRA, ERBB2, ABL1, CDKN1CA, MAP2K6, SRPK3, PDK2A, CKMB, PFKMB, CKMA, NEK6, SI:CH211-22018.4, NEK7, CAMK2B1, PKMA, ALPK3A, PTK2AA, AKT3A, SI:DKEY-8E10.3, MAPKAPK2A, PFKFB4B, SI:DKEY-96F10.1	820	718	18397	1.719	0.149904	0.01476333	0.0145	
6310	GO:000	tricarboxylic acid cycle	Biological	8	0.82	2.80E-04	CS, FH, SUCLA2, MDH2, IDH2, DLST, ACO2, IDH3A	820	30	18397	5.983	0.320254	0.03216517	0.03158
6099	cycle	Process												
GO:004	sarcomerogenesis	Biological	5	0.51	4.11E-04	ZMP:0000000930, TTN.2, TTN.1, TCAP, SMYD1B	820	9	18397	12.464	0.432339	0.04043678	0.0397	
8769	GO:003	skeletal muscle	Biological	5	0.51	4.11E-04	MYOG, CDKN1A, FHL1A, KLHL41B, MYF5	820	9	18397	12.464	0.432339	0.04043678	0.0397
5914	cell differentiation	Process												
GO:005	muscle cell	Biological	9	0.93	5.41E-04	FXR1, CAPZB, MYH7BA, NRAP, TCAP, ACTN3B, NEB, TGFBI, ACTN2B	820	43	18397	4.696	0.52558	0.04969742	0.0488	
5001	development	Process												
GO:004	positive regulation	Biological	6	0.62	6.83E-04	FXR1, FXR2, PCIF1, METTL5, LARP4B, LARP1B	820	17	18397	7.918	0.610285	0.05887603	0.05781	
5727	of translation	Process												
GO:000	cellular calcium	Biological	10	1.03	0.00126	ATP2A2A, RYR1B, RYR1A, ATP2B3B, HOMER1B, TNN1A, ATP2A1, ATP2B2, RYR3, DHRS7CB	820	60	18397	3.739	0.825237	0.10254246	0.10068	
6874	ion homeostasis	Process												

GO:0061061	muscle structure development	Biological Process	6	0.62	0.00153	<i>PDLIM5B, PDLIM3B, LDB3B, HOMER1B, KLHL40B, KLHL40A</i>	820	20	18397	6.731	0.879206	0.11533051	0.11324
GO:0033693	neurofilament bundle assembly	Biological Process	4	0.41	0.00159	<i>SYNM, Si:DKEY-33C12.3, NEFMA, NEFLA</i>	820	6	18397	14.957	0.888421	0.11533051	0.11324
GO:0006470	protein bundle assembly	Biological Process	16	1.65	0.00192	<i>EPIM2A, Si:CH211-223P8.8, PTPN4A, PTPN21, DUSP27, DUSP16, CDC25B, SSH2A, PDP1, PTPRNA, DUSP10, Si:CH211-121A2.2, DUSP22A, Si:CH211-195B15.8, DUSP13A, DUSP22B</i>	820	144	18397	2.493	0.929675	0.13260354	0.1302
GO:0016311	dephosphorylation	Biological Process	15	1.54	0.00243	<i>EPIM2A, Si:CH211-223P8.8, PTPN4A, PTPN21, DUSP27, DUSP16, PTP4A3A, SSH2A, PTPRNA, DUSP10, Si:CH211-121A2.2, DUSP22A, Si:CH211-195B15.8, DUSP13A, DUSP22B</i>	820	133	18397	2.530	0.965226	0.15384033	0.15105
GO:0016567	protein ubiquitination	Biological Process	32	3.29	0.00255	<i>VCP, ANAPC16, KLHL15, ASB5B, PDZRN3B, UBR3, NEDD4L, KLHL13, ASB18, SH3RF1, ASB16, TRIM35-31, CAND2, HERC2, ASB15B, ASB10, DCAF12, CUL3B, Si:CH73-54F23.4, SOCS3B, TRIM55B, FEM1A, KLHL21, ZBTB16A, FBXO32, UBAC1, FBXO31, Si:CH211-120G10.1, ASB2A.1, NEURL2, ASB4, TRIM54</i>	820	406	18397	1.768	0.970353	0.15384033	0.15105
GO:0007623	circadian rhythm	Biological Process	7	0.72	0.00257	<i>NFIL3-6, NROB2A, PER1B, CLOCKA, MITFA, NPAS2, ARNTL2</i>	820	32	18397	4.908	0.97107	0.15384033	0.15105
GO:0007047	heart contraction	Biological Process	11	1.13	0.00284	<i>BAG3, TTN-2, LRRC39, CRYABA, DESMA, DLST, TNNT2A, TCAP, FBXO32, SMYD1B, LIM51</i>	820	80	18397	3.085	0.980211	0.161969	0.15903
GO:0017148	negative regulation of translation	Biological Process	7	0.72	0.00302	<i>FXR1, PAIP2B, FXR2, EIF4EBP1, CAPRIN1A, EIF4EBP3L, YBX1</i>	820	33	18397	4.759	0.984559	0.161969	0.15903
GO:0007015	actin filament organization	Biological Process	16	1.65	0.00305	<i>TMOD1, MYO5AA, TMOD4, TPM3, TPM1, RHOB7B4, MYO16, LMOD3, TMSB, LMOD2B, TPMA, XIRP1, CORO6, BCL2L16, RHOAC, CORO1CA</i>	820	151	18397	2.377	0.985267	0.161969	0.15903
GO:0002922	circadian regulation of gene expression	Biological Process	8	0.82	0.0035	<i>NFIL3-6, BHLHE40, KDM8, CRY2, PER1B, NR1D1, CLOCKA, NPAS2</i>	820	45	18397	3.989	0.992016	0.17843311	0.1752
GO:0003409	negative regulation of MAPK cascade	Biological Process	6	0.62	0.00362	<i>DUSP10, Si:CH211-223P8.8, Si:CH211-121A2.2, Si:CH211-195B15.8, DUSP13A, DUSP16</i>	820	24	18397	5.609	0.993297	0.17843311	0.1752
GO:0006314	phosphocreatine biosynthetic process	Biological Process	4	0.41	0.00416	<i>CKMB, CKMA, CKMT2A, ZGC:172076</i>	820	8	18397	11.218	0.996815	0.18507208	0.18172
GO:0005947	negative regulation of translational initiation	Biological Process	4	0.41	0.00416	<i>PAIP2B, EIF4EBP1, EIF4EBP3L, YBX1</i>	820	8	18397	11.218	0.996815	0.18507208	0.18172
GO:0001243	negative regulation of intrinsic apoptotic signaling pathway	Biological Process	4	0.41	0.00416	<i>MCL1A, MCL1B, BCL2L16, BCL2L1</i>	820	8	18397	11.218	0.996815	0.18507208	0.18172

GO:0030388	fructose 1,6-bisphosphate metabolic process	Biological Process	5	0.51	0.00462	PFKM1B, ALDOAA, ALDOAB, FBP2, ALDOCB	820	16	18397	7.011	0.998316	0.19308118	0.18958
GO:0006937	regulation of muscle contraction	Biological Process	5	0.51	0.00462	TNNI7E, TNNC1B, TNNT3B, TNNT2A, ATP2A1	820	16	18397	7.011	0.998316	0.19308118	0.18958
GO:0070588	calcium ion transmembrane transport	Biological Process	12	1.23	0.00568	ATP2A2A, RYR1B, CACNA1A, RYR1A, CACNA2D2B, ATP2B3B, TRPM4A, ITPR3, ATP2A1, ATP2B2, CACNG1B, RYR3	820	102	18397	2.639	0.999611	0.2302872	0.22611
GO:0051694	pointed-end actin filament capping	Biological Process	4	0.41	0.00604	TMOD1, LMOD2B, TMOD4, LMOD3	820	9	18397	9.971	0.999763	0.23118873	0.227
GO:0055008	cardiac muscle tissue morphogenesis	Biological Process	4	0.41	0.00604	ZMP:0000000930, LRR39, TCAP, FBXO32	820	9	18397	9.971	0.999763	0.23118873	0.227
GO:0055013	cardiac muscle cell development	Biological Process	5	0.51	0.00724	RBFOX2, ZMP:0000000930, RBFOX1L, TCAP, NR2F2	820	18	18397	6.232	0.999955	0.26975824	0.26487
GO:0000240	skeletal muscle thin filament assembly	Biological Process	4	0.41	0.00834	ZMP:0000000930, TCAP, LMOD3, SMYD1B	820	10	18397	8.974	0.99999	0.29486062	0.28952
GO:0006108	malate metabolic process	Biological Process	4	0.41	0.00834	FH, MDH2, ME1, ME3	820	10	18397	8.974	0.99999	0.29486062	0.28952
GO:0030018	Z disc	Cellular Component	26	2.67	4.63E-17	FHL1A, ACTN3B, RYR3, SYNPO21A, SYNPO21B, ZMP:0000000930, BAG3, MYOZ2A, MYOZ2B, NRAP, DESMA, CASQ1B, PDLIM5B, PDLIM3B, NEB, PARVB, ACTN2B, RYR1B, MYOZ3A, RYR1A, LDB3B, MYOZ1A, MYOZ1B, TCAP, TRIM54, LIMS1	827	69	18868	8.597	1.43E-14	1.43E-14	1.37E-14
GO:0016529	sarcoplasmic reticulum	Cellular Component	15	1.54	6.45E-12	KLHL41A, CASQ1B, ITPR3, JPH1A, ATP2A1, JPH1B, RYR3, TRDN, ATP2A2A, RYR1B, JPH2, RYR1A, TMEM38A, KLHL41B, THBS4B	827	30	18868	11.407	1.99E-09	6.78E-10	6.52E-10
GO:0005737	cytoplasm	Cellular Component	278	28.6	6.61E-12	APOBEC2B, UGP2B, LRR34B, PRKAB1A, AGLA, CALCOCO1A, DCAF6, ZFYVE28, NROB2A, ROCK2A, HERC2, EIF2D, DUSP13A, CHAC1, KLHL41B, SMUJ1A, JMJD4, CAVIN4B, KLHL41A, RFX2, ARMC8, RNF123, RUFY3, ULK2, PRKCQ, TLLI2, HPR11, Si:CH211-195B15.8, KLHL40B, KLHL40A, CCNO, UCKL1B, ILRUN, NUMBL, ANAPC16, CAPN3A, DAPK2A, SGIPIA, PTPN4A, NEDD4L, NMD3, APBB2B, TRIM35-31, FXR1, LDHA, FXR2, PCBP4, CMYA5, MYL10, NAA50, SRPK3, MYO5AA, CAMK2B1, PARVB, UBAC1, ALS2B, DAZL, MAPKAPK2A, MYH7L, DUSP22A, ABLIM1A, DUSP22B, FARSB, SYNMM, BTG2, DNAJB6A, SVILA, SETD3, MYLPFB, ACY1, LRR39, RPLP0, PDZRN3B, STON2, SH3RF1, SMG6, SSH2A, MSI2B, GYS1, UCHL1, KIF1B, ZGC:85777, Si:CH73-54F23-4, MYL12.2, PIEKHO1B, CNOT6L, RBFOX1L, MYO18AB, ARG1, MYO18A4, PTGR2, TMSB, EIF4EBP3L, CLOCKA, CACTIN, ACSBG2, KANK2, FH, SIX1B, VCLB, ARNTL2, PAK1, TTN.2, FRZB, DESMA, MAP3K3, Si:CH211-260E23.9, MDH2, Si:CH211-22018.4, STAC3, BBOX1, PKMA, SMC8A, TUBB4B, SMYHC2, RNF146, Si:CH211-120G10.1, IMPDH1A, MSRB3, TACC2, TPI1B, PCMT, LIMS1, FHOD1, PRUNE, CTNND1, UBE3A, IPO4, PAIP2B, GRB14, ULK1B, MID1P1L, ARHGAP10, KCNAB2B, KPNA4, BTBD10A, KPNA3, EGLN1A, EGLN1B, FGF11A, EPM2A, TXNIP, USP4, NEFMA, VASH2, HOMER1B, ZBTB16A, CDC25B, CLIP3, CRTCLB, Si:CH211-253B8.5, ALDH3A2B, MAPKAPK3, TPMA, CRYABA, KPNB3, KANK4, IDI1, VCP, Si:DKEY-	827	4405	18868	1.440	2.03E-09	6.78E-10	6.52E-10

GO:003 M band	Cellular	8	0.82	1.96E-07	SMPX, SPEGB, LRR39, MYOM2A, MYOM1A, MYOM1B, LM0D3, SMYD1B	827	12	18868	15.210	6.05E-05	1.37E-05	1.32E-05
1430	Component											
GO:003 sarcoplasmic reticulum	Cellular	9	0.93	2.23E-07	ATP2A2A, RYR1B, KLHL41A, RYR1A, ATP2A1, TMEM38A, KLHL41B, RYR3, TRDN	827	17	18868	12.079	6.87E-05	1.37E-05	1.32E-05
3017	Component											
GO:001 myosin II complex	Cellular	10	1.03	1.43E-05	SMYHC2, MYL12.2, MYHB, MYO18AB, MYL13, MYO18AA, MYH7B, MYH7L, MYH14, MYLZ3	827	35	18868	6.519	0.004392	7.34E-04	7.05E-04
6460	Component											
GO:000 cytosol	Cellular	77	7.92	2.97E-05	IPO11, UBE3A, MTR, IPO7, DUSP16, LARP1B, ZFYVE28, PSME4B, PSME4A, MAP1LC3A, MID1IP1L, ARHGDI1A, EEF2L2, PGM5, CHAC1, ACY3.1, ZGC:136908, PGM1, Si:DKEY-51E6.1, USP9, ADSL, ADKB, ARG1, PDE4D, ADKA, ACOT12, AMPD1, RBP7B, LARP4B, ALDOAA, ALDOAB, IRS2B, HPR1, Si:CH211-195B15.8, GAPDH, ASPA, FBP2, ZGC:64002, USP13, VCP, FH, AHCY, ANAPC16, RAF1A, AK1, STRIP2, BAG3, G3BP1, CDAB, AAAMP, ZC3H15, NAA50, USP24, GPD1B, RIC1, GPIB, Si:CH211-260E23.9, OSBPL5, PLEKHA5, STAC3, RAD23AA, MITHFR, USP28, ALDOCB, MLT11, RNFI46, ACOT11A, AGBL1, CASTOR2, PPP2R2BB, PFKFB4B, DUSP22A, Si:DKEY-96F10.1, ACO2, TPI1B, LRCH3, DUSP22B	827	1082	18868	1.624	0.009107	0.00130689	0.00126
5829	Component											
GO:000 troponin complex	Cellular	8	0.82	9.52E-05	TNNI1C, TNNI4A, TNNI2A.4, TNNTZE, TNNC1B, TNNI2A.1, TNNT3B, TNNTZA	827	26	18868	7.020	0.028884	0.00366346	0.00352
5861	Component											
GO:004 sarcolemma	Cellular	10	1.03	1.72E-04	POPC3, RYR1B, SGCB, KCNB1, RYR1A, STAC3, DESMA, PGM5, VCLB, RYR3	827	47	18868	4.854	0.051664	0.00589349	0.00566
2383	Component											
GO:001 actin cytoskeleton	Cellular	16	1.65	1.98E-04	SVILA, MYO5AA, ARHGAP32B, VCLB, PARVB, MYO16, SYNPOZLA, SYNPOZLB, MYOZ3A, MYOZ2A, MYOZ2B, MYOZ1A, TPMA, MYOZ1B, ABLIM1A, CORO1CA	827	118	18868	3.094	0.05916	0.00609766	0.00586
5629	Component											
GO:003 myosin filament	Cellular	7	0.72	3.68E-04	SMYHC2, MYHB, MYO18AB, MYO18AA, MYH7B, MYH7L, MYH14	827	23	18868	6.944	0.107093	0.01029566	0.00989
2982	Component											
GO:003 myofibril	Cellular	6	0.62	0.00111	TMOD1, LM0D2B, TM0D4, TNNTZA, TWF2B, LM0D3	827	19	18868	7.205	0.288791	0.02883335	0.02728
0016	Component											
GO:003 filamentous actin	Cellular	6	0.62	0.00225	PDJIM5B, PDJIM3B, LDB3B, EHBP1L1A, EHBP1L1B, SMTNL1	827	22	18868	6.222	0.49967	0.05320843	0.05114
1941	Component											



GO:0016459	myosin complex	Cellular Component	10	1.03	0.00248	SMYHC2, SMYHC3, MYO5AA, MYHB, MYO18AB, MYO18AA, MYH7BA, MYH7L, MYH14, MYO16	827	67	18868	3.405	0.53456	0.05455885	0.05243
GO:0031463	Cu13-RING ubiquitin ligase complex	Cellular Component	6	0.62	0.00675	KLHL15, KLHL21, KLHL13, KLHL40B, KLHL40A, CUL3B	827	28	18868	4.889	0.875746	0.13258904	0.12742
GO:0005856	cytoskeleton	Cellular Component	43	4.42	0.00689	DYRK4, FHOD1, PTPN4A, TUBA8L2, PTPN21, RHOBTB4, HSPB1, LRMP, KRT18A.1, VCLB, CITA, LMOD3, SSH2A, ACTB2, ROCK2A, MID1IP1L, CAPZB, SGCB, DCAF12, KLHL41B, RHOAC, MAP2K6, TMOD1, KLHL41A, ABI2A, TMOD4, CDC135, TPM1, KLHL21, PARVB, TUBB4B, PTK2AA, TMSB, EPB41L3B, EPB41L3A, DNMBP, LMOD2B, TPMA, TWF2B, TACC2, GAPDH, KANK4, CORO1CA	827	645	18868	1.521	0.881016	0.13258904	0.12742
GO:0005911	cell-cell junction	Cellular Component	9	0.93	0.00947	IGSF11, MPP1, EPB41L3B, USPF5B, EPB41L3A, DNMBP, MPP7A, DESMA, LIMSI	827	68	18868	3.020	0.946641	0.17157736	0.16489
GO:0003779	actin binding	Molecular Function	50	5.14	2.02E-13	ABRAB, MYHB, SVILA, ABRAA, SETD3, SSH2A, SYNPO2LA, SYNPO2LB, CAPZB, MYO18AB, TPM3, MYO18AA, PDLIM3B, TPM1, ACTN2B, TMSB, EPB41L3B, DAAM2, MYOZ3A, TPMA, CORO1CA, MICAL2A, TNNT2A, ACTN3B, VCLB, SMTNL, MYOZ2A, NRAP, MYOZ2B, XIRP1, FLNCB, MYH14, FINA, MYO5AA, PDLIM5B, PHACTR3B, NEB, PARVB, MYO16, SMYHC2, SMYHC3, MYH7BA, LDB3B, MYH7L, CAPZA1B, MYOZ1B, ABLIM1A, TWF2B	771	337	17340	3.337	1.41E-10	1.41E-10	1.38E-10
GO:0051015	actin filament binding	Molecular Function	36	3.7	1.19E-09	MYHB, SVILA, FHOD1, ACTN3B, MYOM1A, VCLB, MYOM1B, CAPZB, SPEGB, NRAP, XIRP1, MYH14, FLNA, FLNCB, TMOD1, MYO5AA, MYO18AB, TPM3, MYO18AA, TMOD4, TNNC1B, TPM1, MYOM2A, NEB, MYO16, ACTN2B, SMYHC2, SMYHC3, MYH7BA, CAPZA1B, MYH7L, TPMA, CORO6, ABLIM1A, TWF2B, CORO1CA	771	247	17340	3.278	8.30E-07	4.15E-07	4.08E-07
GO:0008138	protein tyrosine/serine/threonine phosphatase activity	Molecular Function	12	1.23	3.50E-05	EPIM2A, DUSP10, Si:CH211-223P8.8, Si:CH211-121A2.2, DUSP22A, Si:CH211-195B15.8, DUSP13A, DUSP27, DUSP16, DUSP22B, PTP4A3A, SSH2A	771	57	17340	4.735	0.024053	0.00811553	0.00798
GO:0005523	tropomyosin binding	Molecular Function	7	0.72	6.11E-05	TMOD1, LMOD2B, TNNT2E, TMOD4, TNNT3B, TNNT2A, LMOD3	771	17	17340	9.261	0.041562	0.01061236	0.01043
GO:0004879	RNA polymerase II transcription factor activity, ligand-activated	Molecular Function	13	1.34	1.07E-04	ESR2A, RXRAA, RARAA, PPARDB, RORC, RXRGB, NR1D1, NR2F2, NR2F5, NR2F6B, NR2F1A, RORAA, ESRRA	771	75	17340	3.898	0.071924	0.01204106	0.01183
GO:0031433	sequence-specific DNA binding	Molecular Function	5	0.51	1.21E-04	MYOZ3A, MYOZ2A, MYOZ2B, MYOZ1A, MYOZ1B	771	7	17340	16.064	0.080838	0.01204106	0.01183
GO:0051373	FATZ binding	Molecular Function	5	0.51	1.21E-04	MYOZ3A, MYOZ2A, MYOZ2B, MYOZ1A, MYOZ1B	771	7	17340	16.064	0.080838	0.01204106	0.01183
GO:0003824	catalytic activity	Molecular Function	35	3.6	1.67E-04	APOBEC2B, FH, TKTB, AGLA, CKMT2A, GLULB, ACACB, HADHAA, LDHA, HYAL3, ENPP6, CDAB, PFKMB, CKMB, ADSL, PYGMA, CKMA, MDH2, MOCOS, PGAM2, ZGC:172076, PKMA, ALDOAA, ALDOAB, ALDOCB, GOT2B, SUCLA2, IMPDH1A, PCYT1BA, GOT2A, PFKFB4B, Si:DKEY-96F10.1, DUS1L, TPI1B, BCAT1	771	393	17340	2.003	0.109314	0.01446922	0.01422

GO:001 6301	kinase activity	Molecular Function	54	5.56	2.00E-04	771	718	17340	1.691	0.129839	0.01545151	0.01518
GO:001 6791	phosphatase activity	Molecular Function	20	2.06	2.54E-04	771	173	17340	2.600	0.16193	0.01766316	0.01736
GO:000 4721	phosphoprotein phosphatase activity	Molecular Function	17	1.75	4.52E-04	771	139	17340	2.751	0.269638	0.02653149	0.02607
GO:000 3700	transcription factor activity, sequence-specific DNA binding	Molecular Function	46	4.73	4.58E-04	771	602	17340	1.719	0.272725	0.02653149	0.02607
GO:001 6740	transferase activity	Molecular Function	105	10.8	0.00104	771	1743	17340	1.355	0.513478	0.05539232	0.05444
GO:000 4725	protein tyrosine phosphatase activity	Molecular Function	15	1.54	0.00178	771	129	17340	2.615	0.709717	0.08827145	0.08675
GO:000 0978	RNA polymerase II core promoter proximal region sequence-specific DNA binding	Molecular Function	70	7.2	0.00202	771	1095	17340	1.438	0.754295	0.09348039	0.09187
GO:000 4111	creatine kinase activity	Molecular Function	4	0.41	0.00413	771	8	17340	11.245	0.94366	0.17940017	0.1763
GO:005 1019	mitogen-activated protein kinase binding	Molecular Function	5	0.51	0.00578	771	17	17340	6.615	0.982184	0.23623734	0.23216

## Supplementary Table 5

**Supplementary Table 5** 19 Significantly enriched Gene Ontology terms with Fisher's exact test p-value < 0.05 among genes with increasing expression relative to EOD duration (Group 5 and 6).

Term	GO terms	Category	Count	%	P-value	Genes	List	Pop	Pop	Fold	Bonfer	Benjam	FDR
							Total	Hits	Total	Enrichment	roni	ini	
GO:0005975	carbohydrate metabolic process	Biological Process	11	4.564	7.83E-05	CHST7, GNPDA2, MAN2B2, B3GAT3, GLB1, RPE, Si:DKEY-199F5.8, SPATA20, GUSB, YDJC, HK2	205	199	18397	4.961	0.045	0.046	0.046
GO:0048675	axon extension	Biological	4	1.66	0.010509	IST1, UBAP1, PLXNB1B, GRNB	205	41	18397	8.755	0.998	1.000	1.000
GO:0006914	autophagy	Biological Process	5	2.075	0.013656	BECN1, GABARAPL2, PLEKHM1, WIPI1, MCOLN1A	205	83	18397	5.406	1.000	1.000	1.000
GO:0007032	endosome organization	Biological	3	1.245	0.022373	PLEKHF2, IST1, PI4K2A	205	21	18397	12.820	1.000	1.000	1.000
GO:0046854	phosphatidylinositol phosphorylation	Biological Process	4	1.66	0.03282	PIP5K1CA, IMPA1, PIK3CG, PI4K2A	205	63	18397	5.698	1.000	1.000	1.000
GO:0006811	ion transport	Biological Process	13	5.394	0.041416	SLC24A2, SLC12A4, SLC10A7, SLC31A2, KCTD12.1, KCN12A, SFXN5B, ATP6V0A1A, TCN2, ATP6V1D, PANX3, MCU, ATP6V1F	205	614	18397	1.900	1.000	1.000	1.000
GO:0006665	sphingolipid metabolic process	Biological	3	1.245	0.043439	ARV1, PSAP, SFTPB	205	30	18397	8.974	1.000	1.000	1.000
GO:0005764	lysosome	Cellular	11	4.564	3.03E-05	ASAH1B, MAN2B2, BRI3, CTSSA, SMPD1, VPS41, PSAP, PLEKHM1, MCOLN1A, SFTPB, GUSB	201	186	18868	5.551	0.005	0.005	0.005
GO:0005794	Golgi apparatus	Cellular	18	7.469	3.93E-04	ZDHHC16B, BECN1, SLC10A7, YIPF3, B3GAT3, CHPF2, ENTPD6, SYAP1, SCOCA, FUT9A, ARV1, MGAT1B, TRAPPC6B, VPS41, Si:DKEY-199F5.8, EXTL3, ZGC:162698, PI4K2A	201	628	18868	2.691	0.065	0.034	0.034
GO:0005768	endosome	Cellular	11	4.564	6.34E-04	PLEKHF2, BECN1, RAB5B, RAB40C, VPS41, PLEKHM1, UBAP1, FLOT1B, COMMD1, LAMTOR3, PI4K2A	201	270	18868	3.824	0.103	0.036	0.036
GO:0005829	cytosol	Cellular	21	8.714	0.011107	BECN1, GABARAPL2, USP7, RIC1, IRS1, RPE, GSR, DNAJB1A, CST14A.2, PPM1AA, WIPI1, SCOCA, PRDX6, HK2, RSPH1, EIF5, UBAP1, OSBP1A, RAN, ZGC:162698, SPG21	201	1082	18868	1.822	0.854	0.478	0.475
GO:0005773	vacuole	Cellular	3	1.245	0.018738	TMEM138, GLB1, LGMN	201	20	18868	14.081	0.961	0.596	0.592
GO:0031410	cytoplasmic vesicle	Cellular	8	3.32	0.020775	RAC3A, BECN1, VMA21, TBC1D7, VPS41, FLOT1B, RHOCB, PI4K2A	201	259	18868	2.899	0.973	0.596	0.592
GO:0005581	collagen trimer	Cellular	4	1.66	0.02471	COL8A1A, COL12A1A, COL2A1B, COL4A5	201	59	18868	6.364	0.986	0.607	0.604
GO:0005783	endoplasmic reticulum	Cellular	15	6.224	0.032958	ZDHHC16B, PLEKHF2, BECN1, SLC10A7, TAPBP1, HSPBP1, AGPAT2, CYP51, ARV1, ZFYVE27, SPTLC1, TRAPPC6B, VMA21, FKBP9, EXTL3	201	763	18868	1.845	0.997	0.709	0.704
GO:0004185	serine-type carboxypeptidase activity	Molecular Function	3	1.245	6.00E-04	CTSA, SCPEP1, Si:CH211-122F10.4	176	4	17340	73.892	0.166	0.182	0.182
GO:0004180	carboxypeptidase activity	Molecular Function	3	1.245	0.022297	CTSA, SCPEP1, Si:CH211-122F10.4	176	23	17340	12.851	0.999	1.000	1.000
GO:0046961	proton-transporting ATPase activity, rotational mechanism	Molecular Function	3	1.245	0.026095	ATP6V0A1A, ATP6V1D, ATP6V1F	176	25	17340	11.823	1.000	1.000	1.000
GO:0016798	hydrolase activity, acting on glycosyl bonds	Molecular Function	4	1.66	0.047208	MAN2B2, GLB1, SMPD1, GUSB	176	80	17340	4.926	1.000	1.000	1.000

## Supplementary Table 6

**Supplementary Table 6** 41 Significantly enriched Gene Ontology terms with Fisher's exact test p-value <0.05 among genes with decreasing expression relative to EOD duration (Group 3).

Term	GO terms	Category	Count	%	P-value	Genes	List Total	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0007411	axon guidance	Biological Process	12	3.72	8.12E-05	SEMA3AB, CNTN1A, SEMA6A, SEMA4D, LAMA1, NPTNB, EPHA4A, SEMA4C, EFN2A, NRCAMA, TRIOA, DPYSL2B	271	182	18397 4.476	0.055	0.057	0.057
GO:0030198	extracellular matrix organization	Biological Process	9	2.79	7.83E-04	ADAMTS5, COL5A3A, CCDC80, ADAMTS14, COL1A1A, COL2A1B, ADAMTS17, ADAMTS9, ADAMTS15A	271	133	18397 4.594	0.423	0.275	0.274
GO:0006486	protein glycosylation	Biological Process	9	2.79	0.0012	B3GAT1A, GALNT7, NUS1, GALNT14, AARS2, STS1A5, FUT8A, LARGE2, ST3GAL2	271	142	18397 4.303	0.569	0.280	0.280
GO:0031103	axon regeneration	Biological Process	4	1.24	0.00183	B3GAT1A, SEMA4D, LINGO1A, DPYSL2B	271	17	18397 15.973	0.723	0.295	0.294
GO:0007155	cell adhesion	Biological Process	16	4.95	0.0021	DCHS1B, PCDH15B, EGFL6, LAMA1, PCDH19, CCN2B, CNTN1A, VCANB, EPHA4A, ITGA10, NPTNB, EFN2A, NCAM1A, ITGB8, FBN2B, BCAR1	271	437	18397 2.486	0.771	0.295	0.294
GO:0016310	phosphorylation	Biological Process	21	6.5	0.00475	UCK2A, CHK8, MARK3B, DGKB, PFKFB2B, EIF2AK3, DGKZA, NUCKS1A, CDK7, CDK5, EPHB4A, ETNK2, EPHA4A, GRK6, CXBB, Si:CH211-243120.2, MAP3K8, MAPK14A, TRIOA, PIK3R3B, CAMK1DA	271	718	18397 1.986	0.965	0.556	0.555
GO:0008045	motor neuron axon	Biological Process	5	1.55	0.00807	SEMA3AB, LAMA1, BICD1A, NOTUM2, NTN2	271	54	18397 6.286	0.997	0.810	0.809
GO:0048843	negative regulation of axon extension involved in axon	Biological Process	4	1.24	0.01447	SEMA3AB, SEMA6A, SEMA4D, SEMA4C	271	35	18397 7.758	1.000	1.000	1.000
GO:0007420	brain development	Biological Process	9	2.79	0.02468	ADGRL2A, SCGN, CNTN1A, SEMA4D, LAMA1, TAOK2A, NRCAMA, PCDH19, DPYSL2B	271	238	18397 2.567	1.000	1.000	1.000
GO:0050919	negative chemotaxis	Biological Process	4	1.24	0.02506	SEMA3AB, SEMA6A, SEMA4D, SEMA4C	271	43	18397 6.315	1.000	1.000	1.000
GO:0006468	protein phosphorylation	Biological Process	19	5.88	0.02577	ADAM10B, MARK3B, TAOK2A, EIF2AK3, GUCY2F, CDK7, BRSK2B, CDK5, EPHB4A, EPHA4A, GRK6, MAP3K8, MAPK14A, TRIOA, FAM20CB, NRBP2B, CAMK1DA, Si:CH211-1111.3, MARK1	271	741	18397 1.741	1.000	1.000	1.000
GO:0035475	angioblast cell migration involved in selective angioblast	Biological Process	2	0.62	0.02914	EPHB4A, EFN2A	271	2	18397 67.886	1.000	1.000	1.000
GO:0033334	fin morphogenesis	Biological Process	3	0.93	0.03114	FRAS1, COL1A1A, COL2A1B	271	19	18397 10.719	1.000	1.000	1.000
GO:0090630	activation of GTPase activity	Biological Process	5	1.55	0.03384	ARHGAP22, Si:CH211-288D18.1, Si:DKEY-191M6.4, SIPA1L1, AGAP3	271	83	18397 4.089	1.000	1.000	1.000
GO:0048013	ephrin receptor signaling pathway	Biological Process	3	0.93	0.03751	EPHB4A, EFN2A, ANKS1A1B	271	21	18397 9.698	1.000	1.000	1.000
GO:0071526	semaphorin-plexin signaling pathway	Biological Process	4	1.24	0.03885	SEMA3AB, SEMA6A, SEMA4D, SEMA4C	271	51	18397 5.324	1.000	1.000	1.000
GO:0003404	optic vesicle	Biological Process	2	0.62	0.04339	EPHA4A, EFN2A	271	3	18397 45.257	1.000	1.000	1.000
GO:0001756	somitogenesis	Biological Process	5	1.55	0.04654	CCDC80, MEF2AA, EFN2A, MAPK14A, DLC	271	92	18397 3.689	1.000	1.000	1.000
GO:0030903	notochord	Biological Process	4	1.24	0.04692	COL5A3A, EGFL6, LAMA1, COL2A1B	271	55	18397 4.937	1.000	1.000	1.000
GO:0031012	extracellular matrix	Cellular Component	13	4.02	1.29E-04	COLEC12, COL5A3A, CCN2B, ADAMTS5, VCANB, ADAMTS14, COL1A1A, ADAMTS17, COL2A1B, FBN2B, ADAMTS9, LINGO1A, ADAMTS15A	287	218	18868 3.920	0.023	0.024	0.024

GO:0005794	Golgi apparatus	Cellular Component	21	6.5	0.00151	GALNT7, ARHGAP32A, ADAM10B, GALNT14, AARS2, CLASP1A, SACM11B, FUT8A, ZDHHC23B, B4GALNT3A, ZDHHC14, B3GAT1A, ST8SIA5, ZGC:162200, FAM20CB, BICD1A, LARGE2, GRINAA, ERGIC1, ST3GAL2, PLA2G4AB	287	628	18868	2.198	0.241	0.138	0.138
GO:0016020	membrane	Cellular Component	129	39.9	0.00816	ADGRG2A, GALNT14, IFITM1, DGKB, RHOT1B, OLFCS1, XYLT1, ACSL4A, TMEM263, FADS2, SI:CH211-1E14.1, TNFSF11, EHD1B, SI:DKY-174M14.3, LARGE2, INPP4AB, SYPL2B, ADAM10B, ENTPD2A-1, SEMA6A, NUP210, ATP6AP1A, GPM2L, FUT8A, TRIM101, SLC5A9, ADAM15, ADGRB2, EPHA4A, MADD, TMCC1B, FAM20CB, GRAMD1BB, OSBP13B, DIPK1B, WAI2, AARS2, RPN2, SDC3, MTMR9, SACM11B, PCDH19, RNF145B, HSD11B2, TMEM248, VPS11, ST8SIA5, GSG1L, SCN3B, ST3GAL2, LINGO1A, NUS1, TMEM132E, ANKRD22, TRPV1, SLC25A25B, FLRT1B, EPHB4A, NPTNB, LRFN1, ITGA10, SPIRE2, TSPAN5A, NCAM1A, SDK1A, LRCH4, TMEM19, MCTP1A, MACF1A, PIGS, KCNK6, TRPC6A, NRCAMA, VSIG10, B4GALNT3A, SPRED2A, SI:CH1073-291C23.2, DGKZA, CATIP, SI:DKY-112M2.1, CPT2, EFNB2A, ITGB8, GALST4, TMEM119B, DCHS1B, PCDH15B, EGFL6, ANO6, MMEL1, SLC39A14, ZDHHC14, RRP1A, B3GAT1A, KCNK4A, RAP2B, ECRG4A, TMEM218, BIRC6, DLC, NAT8L, GRINAA, TNFRSF21, ERGIC1, COLEC12, GRAMD1C, LAMA1, PIP4P1A, GUCY2F, ADCY7, PPP1R3AA, IGSF21A, CNTN1A, SI:CH211-278J3.3, GDDP4A, KCNQ5A, KCNQ5B, GALNT7, SLC35A1, SEMA4C, ZDHHC23B, MGAT4C, SPCS2, JAGN1B, FRAS1, GDDP5B, EPPK1, SI:CH73-267C23.10, GPR146	287	7112	18868	1.192	0.777	0.488	0.488
GO:0005576	extracellular region	Cellular Component	27	8.36	0.01066	LAMA1, IGFBP6B, CXCL19, OLFM3A, FGF2, HTRA1B, NTN2, CXCL18A.1, SEMA3AB, ADAMTS5, GLI2, ADAMTS14, COL1A1A, ADAMTS17, ENPP2, FBN2B, NOTUM2, ADAMTS9, MARK1, ADAMTS15A, COLSA3A, CCN2B, VCANB, CDC80, DKK3A, ECRG4A, COL2A1B	287	1059	18868	1.676	0.859	0.488	0.488
GO:0043231	intracellular membrane-bounded	Cellular Component	11	3.41	0.02417	EPN3A, HSD11B2, SPCS2, PHILP1, HIP1RB, PDXDC1, DHRS9, EHD1B, NAT8L, RHOBTB1, OSBP13B	287	319	18868	2.267	0.989	0.885	0.885
GO:0000139	Golgi membrane	Cellular Component	11	3.41	0.03016	B3GAT1A, GALNT7, SLC35A1, GALNT14, AARS2, XYLT1, SACM11B, ZDHHC23B, LARGE2, ERGIC1, ZDHHC14	287	331	18868	2.185	0.996	0.886	0.886
GO:0015629	actin cytoskeleton	Cellular	6	1.86	0.0339	MTSS1LA, ARHGAP32A, CATIP, HAX1, MYO5B, MACF1A	287	118	18868	3.343	0.998	0.886	0.886
GO:0005930	axoneme	Cellular	4	1.24	0.0421	HYDIN, CFAP36, CFAP206, DNALI1	287	51	18868	5.156	1.000	0.911	0.911
GO:0005938	cell cortex	Cellular	5	1.55	0.04482	ARHGAP32A, SPIRE2, GPM2L, RHOBTB1, MACF1A	287	88	18868	3.735	1.000	0.911	0.911
GO:0005509	calcium ion binding	Molecular Function	24	7.43	4.28E-04	DCHS1B, SNED1, PCDH15B, DIPK1B, CALM1A, EGFL6, DGKB, RHOT1B, PDCD6, ACTN1, PCDH19, SLC25A25B, EDIL3A, SCGN, VCANB, PPP3R1B, EFHD2, ENPP2, DLC, FBN2B, EHD1B, MCTP1A, MACF1A, PLA2G4AB	250	739	17340	2.253	0.139	0.150	0.150
GO:0004222	metalloendopeptidase activity	Molecular Function	8	2.48	0.00191	ADAMTS5, ADAM10B, ADAM15, ADAMTS14, ADAMTS17, MMEL1, ADAMTS9, ADAMTS15A	250	122	17340	4.548	0.488	0.323	0.323
GO:0016740	transferase activity	Molecular Function	40	12.4	0.00346	UBE2NB, UGP2B, UCK2A, GALNT14, AARS2, DGKB, PFKFB2B, XYLT1, B4GALNT3A, DGKZA, SI:CH211-278J3.3, CPT2, GRK6, ST8SIA5, SI:CH211-243J20.2, MAP3K8, TRIOA, LARGE2, PIK3R3B, ST3GAL2, GALNT7, NUS1, CHKB, MARK3B, EIF2AK3, FUT8A, ZDHHC23B, ZDHHC14, MGAT4C, B3GAT1A, NUCKS1A, CDK7, CDK5, ETKN2, EPHB4A, EPHA4A, CKBB, MAPK14A, NAT8L, CAMK1DA	250	1743	17340	1.592	0.703	0.323	0.323

GO:0016301	kinase activity	Molecular Function	21	6.5	0.0037	UCK2A, CHK8, MARK3B, DGKB, PFKFB2B, EIF2AK3, DGKZA, NUCKS1A, CDK7, CDK5, EPHB4A, ETNK2, EPHA4A, GRK6, CKBB, Si:CH211-243J20.2, MAP3K8, MAPK14A, TRIOA, PIK3R3B, CAMK1DA	250	718	17340	2.029	0.726	0.323	0.323
GO:0030215	semaphorin receptor	Molecular Function	4	1.24	0.01261	SEMA3AB, SEMA6A, SEMA4D, SEMA4C	250	34	17340	8.160	0.988	0.779	0.779
GO:0045499	chemorepellent	Molecular Function	4	1.24	0.01473	SEMA3AB, SEMA6A, SEMA4D, SEMA4C	250	36	17340	7.707	0.994	0.779	0.779
GO:0016757	transferase activity, transferring glycosyl	Molecular Function	10	3.1	0.01935	B3GAT1A, GALNT7, MGAT4C, GALNT14, AARS2, ST8SIA5, XYL1, FUT8A, LARGE2, ST3GAL2	250	278	17340	2.495	0.999	0.779	0.779
GO:0005178	integrin binding	Molecular Function	5	1.55	0.01995	CCN2B, EGFL6, ADAM15, ITGA10, ITGB8	250	72	17340	4.817	0.999	0.779	0.779
GO:0050321	tau-protein kinase	Molecular Function	3	0.93	0.02159	BRSK2B, MARK3B, MARK1	250	16	17340	13.005	1.000	0.779	0.779
GO:0005524	ATP binding	Molecular Function	35	10.8	0.02225	UBE2NB, UCK2A, AARS2, DGKB, PFKFB2B, TAOK2A, GUCY2F, SMC4, ADCY7, DGKZA, SMCHD1, Si:CH211-257P13.3, GRK6, DHX15, Si:CH211-243J20.2, MAP3K8, TRIOA, EHD1B, NRBP2B, Si:CH211-J111.3, MARK1, ACTR3, MARK3B, ENTPD2A-1, STAR9, EIF2AK3, CDK7, BRSK2B, CDK5, EPHA4A, EPHA4A, MYO5B, CKBB, MAPK14A, CAMK1DA	250	1662	17340	1.461	1.000	0.779	0.779
GO:0016758	transferase activity, transferring hexosyl	Molecular Function	5	1.55	0.0328	GALNT7, GALNT14, AARS2, LARGE2, B4GALNT3A	250	84	17340	4.129	1.000	1.000	1.000
GO:0005085	guanyl-nucleotide exchange factor	Molecular Function	8	2.48	0.03611	PREX2, BCAR3, RABGEF1, CCDC88C, MADD, ARHGGEF9B, TRIOA, RAPGEF11	250	215	17340	2.581	1.000	1.000	1.000
GO:0005201	extracellular matrix structural constituent	Molecular Function	5	1.55	0.045	COL5A3A, LAMA1, COL1A1A, COL2A1B, FBN2B	250	93	17340	3.729	1.000	1.000	1.000

## Supplementary Table 7

Supplementary Table 7 Imbalanced expressed alleles and their *C. compressirostris* allele proportion for all five replicates of any hybrid cohort.

Hybrid cohort	Tissue	SNPs ID	Gene ID in Annotation	Proportion 1	Proportion 2	Proportion 3	Proportion 4	Proportion 5	Average Proportion (95% Confidence Limits)	Gene	Highlights of Predicted Function	Gene Description
com x rhy	EO	1665681	maker-ptg0002671-snap-gene-2.67-mRNA-1	0.64	0.63	0.64	0.7	0.67	0.66 (0.62-0.369)	<i>ANKRD12</i>	ankyrin repeats-containing cofactor	ankyrin repeat domain 12
com x rhy	EO	681658	maker-ptg0000821-snap-gene-20.7-mRNA-1	0.64	0.61	0.74	0.68	0.73	0.68 (0.61-0.75)	<i>ARL13b</i>	cilium-specific protein	ADP ribosylation factor like GTPase 13B
com x rhy	EO	2982031	maker-ptg0005981-snap-gene-2.10-mRNA-1	0.65	0.79	0.81	0.68	0.67	0.72 (0.63-0.81)	<i>CDH15</i>	calcium-dependent cell adhesion protein	cadherin 15
com x rhy	EO	1098949	maker-ptg0001601-snap-gene-11.44-mRNA-1	0.86	0.82	0.79	0.81	0.79	0.82 (0.78-0.85)	<i>CHRD</i>	opening of an ion-conducting channel across the plasma membrane.	cholinergic receptor nicotinic delta subunit
com x rhy	EO	4997264	maker-ptg0015361-augustus-gene-4.34-mRNA-1	0.61	0.61	0.79	0.6	0.64	0.65 (0.60-0.71)	<i>COL6a3</i>	cell-binding protein	collagen type VI alpha 3 chain
com x rhy	EO	4997265	maker-ptg0015361-augustus-gene-4.34-mRNA-1	0.61	0.62	0.8	0.6	0.64				
com x rhy	EO	4531902	maker-ptg0012361-augustus-gene-26.64-mRNA-1	0.71	0.75	0.79	0.74	0.66	0.73 (0.67-0.79)	<i>CRELD1</i>	epidermal growth factor-related proteins	cysteine rich with EGF like domains 1
com x rhy	EO	4531904	maker-ptg0012361-augustus-gene-26.64-mRNA-1	0.81	0.6	0.65	0.89	0.7				
com x rhy	EO	1396998	maker-ptg0002161-augustus-gene-4.4-mRNA-1	0.67	0.6	0.64	0.61	0.6	0.62 (0.59-0.66)	<i>DAG1</i>	laminin and basement membrane assembly	dystroglycan 1
com x rhy	EO	4392819	snap_masked-ptg0011651-processed-gene-0.145-mRNA-1	0.65	0.74	0.67	0.77	0.74	0.71 (0.65-0.78)	<i>DCAF6</i>	ligand-dependent coactivator of nuclear receptors	DDB1 and CUL4 associated factor 6
com x rhy	EO	837824	maker-ptg0001101-snap-gene-11.31-mRNA-1	0.67	0.61	0.71	0.61	0.61	0.64 (0.58-0.70)	<i>DST</i>	cytoskeletal linker protein	dystonin
com x rhy	EO	3404159	maker-ptg0007291-snap-gene-5.79-mRNA-1	0.76	0.8	0.7	0.69	0.84	0.74 (0.69-0.79)	<i>ENPP2</i>	phosphodiesterase	ectonucleotide pyrophosphatase/phosphodiesterase 2
com x rhy	EO	3404158	maker-ptg0007291-snap-gene-5.79-mRNA-1	0.73	0.67	0.68	0.7	0.84				
com x rhy	EO	3813376	maker-ptg0008761-snap-gene-1.7-mRNA-1	0.66	0.67	0.71	0.63	0.6	0.65 (0.60-0.71)	<i>HOXC11a</i>	multicellular organism development and regulation of transcription	homeobox C11a
com x rhy	EO	1651431	maker-ptg0002651-est_gff_est2genome-gene-6.33-mRNA-1	0.2	0.18	0.13	0.14	0.12	0.16 (0.11-0.20)	<i>KCNJ2</i>	allow potassium to flow into a cell rather than out of a cell, probably participates in establishing action potential waveform	inward rectifier potassium channel 2
com x rhy	EO	768530	snap_masked-ptg0001001-processed-gene-4.104-mRNA-1	0.81	0.69	0.77	0.88	0.77				
com x rhy	EO	768526	snap_masked-ptg0001001-processed-gene-4.104-mRNA-1	0.62	0.62	0.6	0.67	0.68	0.71 (0.64-0.78)	<i>OBSCN</i>	structural component of striated muscles which plays a role in myofibrillogenesis	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF

com x rhy	EO	1624861	maker-ptg0002541-augustus-gene-4.109-mRNA-1	0.69	0.9	0.68	0.9	0.73	0.78 (0.64-0.92)	PALMD	regulation of cell shape	palmdelphin
com x rhy	EO	2766503	maker-ptg0005481-snap-gene-0.90-mRNA-1	0.62	0.78	0.63	0.67	0.76	0.69 (0.60-0.78)	PNKD	regulation of myofibrillogenesis	paroxysmal nonkinetogenic dyskinesia
com x rhy	EO	4446575	maker-ptg0011881-snap-gene-6.4-mRNA-1	0.66	0.63	0.61	0.69	0.69	0.66 (0.64-0.68)	SCN4aa	voltage-gated sodium channel	sodium channel protein type 4 subunit alpha A
com x rhy	EO	4446574	maker-ptg0011881-snap-gene-6.4-mRNA-1	0.65	0.68	0.67	0.66	0.63				
com x tsh	EO	2670824	maker-ptg0005091-augustus-gene-3.41-mRNA-1	0.91	0.85	0.78	0.92	0.86	0.86 (0.79-0.93)	TSPAN7b	integral component of plasma membrane	tetraspanin 7b
com x tsh	EO	5058039	maker-ptg0015861-snap-gene-1.60-mRNA-1	0.94	0.71	0.82	0.9	0.9	0.84 (0.77-0.91)	UACA	modulates isoactin dynamics to regulate the morphological alterations required for cell growth and motility	uveal autoantigen with coiled-coil domains and ankyrin repeats
com x tsh	EO	5058040	maker-ptg0015861-snap-gene-1.60-mRNA-1	0.91	0.68	0.77	0.9	0.87				
com x rhy	SM	3177368	maker-ptg0009591-snap-gene-7.19-mRNA-1	0.72	0.6	0.67	0.67	0.68	0.67 (0.61-0.72)	PLPP3	membrane glycoprotein at the cell plasma membrane	phospholipid phosphatase 3
com x rhy	SM	305826	maker-ptg0000511-snap-gene-6.91-mRNA-1	0.73	0.67	0.61	0.68	0.64	0.67 (0.61-0.72)	HSP90aa1	inducible molecular chaperone	heat shock protein 90 alpha family class a member 1
com x rhy	SM	3786101	maker-ptg0013941-snap-gene-3.24-mRNA-1	0.66	0.7	0.65	0.8	0.6	0.68 (0.59-0.78)	MYO18a	unconventional myosin	myosin XVIIIa
com x rhy	SM	4016401	maker-ptg0016191-snap-gene-1.135-mRNA-1	0.65	0.75	0.79	0.64	0.72	0.71 (0.63-0.79)	TRIM63	muscle-specific rING finger protein	E3 ubiquitin-protein ligase TRIM63
com x rhy	SM	862477	maker-ptg0001601-snap-gene-11.44-mRNA-1	0.8	0.8	0.7	0.69	0.83	0.76 (0.68-0.84)	CHRNA	opening of an ion-conducting channel across the plasma membrane	cholinergic receptor nicotinic delta subunit
com x rhy	SM	606006	snap_masked-ptg0001001-processed-gene-4.104-mRNA-0.84	0.7	0.83	0.84	0.84	0.75	0.79 (0.71-0.87)	OBSCN	structural component of striated muscles which plays a role in myofibrillogenesis	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF
com x rhy	SM	3426233	maker-ptg0011561-augustus-gene-8.0-mRNA-1	0.91	0.9	0.92	0.95	0.92	0.92 (0.90-0.94)	CASQ1a	skeletal muscle specific member of the calsequestrin protein	calsequestrin-1a
com x tsh	SM	1814167	maker-ptg0004131-snap-gene-5.117-mRNA-1	0.77	0.86	0.73	0.78	0.81	0.79 (0.73-0.85)	XIRP1	protect actin filaments during depolymerization	xin actin-binding repeat-containing protein 1



Supplementary Table 8

**Supplementary Table 8** Sequence information of *KCNJ2* transcripts in all pure-bred individuals

com 1	<p>ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACCTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG  GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGTCTTCCAAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC  AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC  AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC  TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGTCCCAG  CAGTGTGTCTCCAATGTCAACAGCTTTCATGGCAGCCTTCTTCTCTGTGGAGACGCAGACCACTATTGGCTATGGTTACCACCATGT  GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTC  TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCTGGTGTGTTAGCCACAACGCTACAATAGCAATGCGGGACGGCAAGCTA  TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCAGACC  GCCGAGGGGAGTTTATCCCCCTAGACCAGTATGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTGTTTCCCCCA  TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT  CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG  CTTCGAGCCTGTACTCTTTGAGGAGAAGAATACTACAAGGTCGACTACTCTCATTTCACAAAACCTACGAGGTGCCGAGCACTCCG  CTATGTAGTGCAGGGAGCTTGCTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGGCTATGAGAATGAAGTGGCGATGATG  GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT  CATGACGTTTATAGACGAGAATCTGAGATTTGA</p>
com 2	<p>ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACCTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG  GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGTCTTCCAAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC  AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC  AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC  TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGTCCCAG  CAGTGTGTCTCCAATGTCAACAGCTTTCATGGCAGCCTTCTTCTCTGTGGAGACGCAGACCACTATTGGCTATGGTTACCACCATGT  GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTC  TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCTGGTGTGTTAGCCACAACGCTACAATAGCAATGCGGGACGGCAAGCTA  TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCAGACC  GCCGAGGGGAGTTTATCCCCCTAGACCAGTATGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTGTTTCCCCCA  TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT  CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG  CTTCGAGCCTGTACTCTTTGAGGAGAAGAATACTACAAGGTCGACTACTCTCATTTCACAAAACCTACGAGGTGCCGAGCACTCCG  CTATGTAGTGCAGGGAGCTTGCTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGGCTATGAGAATGAAGTGGCGATGATG  GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT  CATGACGTTTATAGACGAGAATCTGAGATTTGA</p>
com 3	<p>ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACCTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG  GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGTCTTCCAAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC  AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC  AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC  TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGTCCCAG  CAGTGTGTCTCCAATGTCAACAGCTTTCATGGCAGCCTTCTTCTCTGTGGAGACGCAGACCACTATTGGCTATGGTTACCACCATGT  GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTC  TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCTGGTGTGTTAGCCACAACGCTACAATAGCAATGCGGGACGGCAAGCTA  TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCAGACC  GCCGAGGGGAGTTTATCCCCCTAGACCAGTATGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTGTTTCCCCCA  TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT  CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG  CTTCGAGCCTGTACTCTTTGAGGAGAAGAATACTACAAGGTCGACTACTCTCATTTCACAAAACCTACGAGGTGCCGAGCACTCCG  CTATGTAGTGCAGGGAGCTTGCTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGGCTATGAGAATGAAGTGGCGATGATG  GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT  CATGACGTTTATAGACGAGAATCTGAGATTTGA</p>
com 4	<p>ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACCTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG  GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGTCTTCCAAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC  AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC  AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC  TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGTCCCAG  CAGTGTGTCTCCAATGTCAACAGCTTTCATGGCAGCCTTCTTCTCTGTGGAGACGCAGACCACTATTGGCTATGGTTACCACCATGT  GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTC  TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCTGGTGTGTTAGCCACAACGCTACAATAGCAATGCGGGACGGCAAGCTA  TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCAGACC  GCCGAGGGGAGTTTATCCCCCTAGACCAGTATGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTGTTTCCCCCA  TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT  CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG  CTTCGAGCCTGTACTCTTTGAGGAGAAGAATACTACAAGGTCGACTACTCTCATTTCACAAAACCTACGAGGTGCCGAGCACTCCG  CTATGTAGTGCAGGGAGCTTGCTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGGCTATGAGAATGAAGTGGCGATGATG  GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT  CATGACGTTTATAGACGAGAATCTGAGATTTGA</p>

com 5	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAAACTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTC AAGAAGGATGGACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGTCTCCAATGTCAACAGCTTCATGGCAGCCTTCTCTTCTGTGGAGACGCAGACCCTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTCA TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGGTTAGCCACAACGCTACAATAGCAATGGCGGACGGCAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCACGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGCTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGACTCCG CTATGTAGTGC GCGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
tsh 1	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAAACTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTC AAGAAGGATGGACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGTCTCCAATGTCAACAGCTTCATGGCAGCCTTCTCTTCTGTGGAGACGCAGACCCTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTCA TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGGTTAGCCACAACGCTACAATAGCAATGGCGGACGGCAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCACGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGCTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGACTCCG CTATGTAGTGC GCGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
tsh 2	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAAACTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTC AAGAAGGATGGACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGTCTCCAATGTCAACAGCTTCATGGCAGCCTTCTCTTCTGTGGAGACGCAGACCCTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTCA TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGGTTAGCCACAACGCTACAATAGCAATGGCGGACGGCAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCACGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGCTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGACTCCG CTATGTAGTGC GCGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
tsh 3	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAAACTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAATGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTC AAGAAGGATGGACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGTCTCCAATGTCAACAGCTTCATGGCAGCCTTCTCTTCTGTGGAGACGCAGACCCTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCGACGCCTTCATCATTGGTGCCGTCA TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGGTTAGCCACAACGCTACAATAGCAATGGCGGACGGCAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCACGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGCTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGACTCCG CTATGTAGTGC GCGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
rhy 1	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAAACTTTTCCAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAGTGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTC AAGAAGGATGGACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGTCTCCAATGTCAACAGCTTCATGGCAGCCTTCTCTTCTGTGGAAACGCAGACCCTATTGGCTATGGTTACCACCATGT

	GACAGAAGAGTGCCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCAACGCCTTCATCATTGGTGCCGTCATG TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGTGTTAGCCACAACGCCTACAATAGCAATGCGGGACGGTAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCGGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCAGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGCACTCCG CTATGTAGTGCAGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTTGCTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
rhy 2	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACTTTTCCAAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAGTGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGCTCCAATGTCAACAGCTTCATGGCAGCCTTCTTCTCTGTGGANACGCAGACCACTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCAACGCCTTCATCATTGGTGCCGTC TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGTGTTAGCCACAACGCCTACAATAGCAATGCGGGACGGTAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCAGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGCACTCCG CTATGTAGTGCAGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTTGCTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
rhy 3	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACTTTTCCAAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAGTGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGCTCCAATGTCAACAGCTTCATGGCAGCCTTCTTCTCTGTGGANACGCAGACCACTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCAACGCCTTCATCATTGGTGCCGTC TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGTGTTAGCCACAACGCCTACAATAGCAATGCGGGACGGTAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCAGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGCACTCCG CTATGTAGTGCAGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTTGCTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
rhy 4	ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACTTTTCCAAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGATG GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC AGCAGTGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCATC AACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTTC TGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCCAG CAGTGTGCTCCAATGTCAACAGCTTCATGGCAGCCTTCTTCTCTGTGGAAACGCAGACCACTATTGGCTATGGTTACCACCATGT GACAGAAGAGTGGCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCAACGCCTTCATCATTGGTGCCGTC TGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGTGTTAGCCACAACGCCTACAATAGCAATGCGGGACGGTAAGCTA TGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGACCACC GCCGAGGGGGAGTTTATCCCCCTAGACCAGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTTGTTTCCCCCA TCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGAGACTGCTGGATTTGAGATTGTGGT CATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACACTG CTTCGAGCCTGTACTCTTTGAGGAGAAGAAGTACTACAAGGTCGACTACTCTCATTTCACAAAAACCTACGAGGTGCCGAGCACTCCG CTATGTAGTGCAGGGAGCTTGTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTTGCTATGAGAATGAAGTGGCGATGATG GACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTCTT CATGACGTTTAGACGAGAATCTGAGATTTGA
rhy 5	"ATGAATGTCCAAAACCTGTTCTCAGAAGGTTCTTCCAAAACTTTTCCAAAGGCAGCAGAAGTGAAGCCCTCATCAGCAGAAGCGAT GGTAGTGTGCGGGCCAGCCGCTACAGCGTCGTGCTCCTCCAAAGTAGATGGCCTCAAGTTGGCCACTGTGGCCGTGTCCAATGGCCAC CAGCAGTGGTGTATGGCAAGGTGAACATGTGGCAGCCGGTGCCATGTCGTTTCGTCGAAGAAGGATGGACACTGCAACGTGCACATCAT CAACATGAGCGAGAAAAGGCCAGCGCTACATAGCCGACATCTTACCACCTGCGTGGACATCCGTTGGCGATGGATGATAATCATCTT CTGCTTGACTTTTGTGCTTTCATGGTTGTTCTTTGGCTATGTGTTCTGGCTGGTGGCCTTCTTCTATGGTGACTTGGGGAATAGCTCCC GCAGTGTGCTCCAATGTCAACAGCTTCATGGCAGCCTTCTTCTCTGTGGANACGCAGACCACTATTGGCTATGGTTACCACCATGT GTGACAGAAGAGTGGCCATCGCTGTCTTTATGGTGGTTTTCCAGTGCATTGTTGGCTGCATCATCAACGCCTTCATCATTGGTGCCG CATGGCCAAGATGGCCAAGCCACGAAGCGCAATGAAACCCCTGGTGTGTTAGCCACAACGCCTACAATAGCAATGCGGGACGGTAAGC TATGCCTGATGTGGCGAGTTGGCAACCTACGCAAAAAGCCACCTGGTGGAGGCCACGTGAGGGCTCAGCTACTCAAGTCCCGACCACC CCGAGGGGGAGTTTATCCCCCTAGACCAGTAGATATTGATGTGGGCTTTGACACTGGCGTAGACCCGGATCTTCTTGTTTCCCCCA CATCACCATTGTCCATGAGATCAACGAGGACAGTCCCTTCTATGATATGAGCAAGCAGGATTTTGAGACTGCTGGATTTGAGATTGTGG GTCATCCTGGAGGGCATGGTAGAAGCCACAGCCATGACAACCCAGTGTGCGAGTTTCTACCTGGCAGGGGAGATCCTCTGGGGACAC

TGCTTCGAGCCTGTA CTCTTTGAGGAGAAGA ACTACTACAAGGTCGACTACTCTCATTTCACAAAACCTACGAGGTGCCGAGCACTC  
CGCTATGTAGTGC GCGGGAGCTTGCTGAAAAGAAGGATAATGAGTCCAGCTCTAACTCTTTTGCTATGAGAATGAAGTGGCGATGA  
TGGACAAAAGAGGAGACGGAGGACAAAAGCGAGTGCAGCAATGATGGGAGCAGTTCACAAAAGGCTTCAGAGTTGGGGCGCAATCTC  
TTCATGACGTTTAGACGAGAATCTGAGATTGA

Supplementary Table 9

**Supplementary Table 9** Prediction of impact of two amino acid substitutions inferred from KCNJ2 transcripts of *com*, *tsh* and *rhy*.

Site(aa)	<i>com</i>	<i>tsh</i>	<i>rhy</i>	Score	Sensitivity	Specificity	Prediction
60	N	N	S	0,223	0,91	0,88	Benign
198	D	D	N	0,983	0,74	0,96	Probably damaging

## 5 Article III

### **Gene and allele specific expression during electric organ ontogeny in African weakly electric fish and their hybrids**

Feng Cheng<sup>1</sup>, Alice B. Dennis<sup>1, 2</sup>, Otto Baumann<sup>3</sup>, Frank Kirschbaum<sup>1, 4</sup>, Marisol Domínguez<sup>1</sup>,  
Ralph Tiedemann<sup>1,\*</sup>

1 Unit of Evolutionary Biology and Systematic Zoology, Institute of Biochemistry and Biology,  
University of Potsdam, Potsdam, Germany

2 Laboratory of Adaptive Evolution and Genomics, URBE, Institute of Life, Earth & Environment,  
University of Namur, Namur, Belgium

3 Department of Animal Physiology, Institute of Biochemistry and Biology, University of Potsdam,  
Potsdam, Germany

4 Department of Crop and Animal Science, Faculty of Life Sciences, Humboldt University, Berlin,  
Germany

#### **Abstract**

Hybridization can act as a catalyst for rapid phenotypic evolution by introducing novel allelic combinations, which can affect hybrid phenotype through changes in gene expression. The African weakly electric fish use their muscle-derived electric organ to produce an electric organ discharge (EOD) for electrocommunication and electrolocation. The EOD in the genus *Campylomormyrus* and cross-species hybrids is usually species/cross-specific and varies during ontogeny. We

compared the gene expression patterns and allele specific expression between juvenile and adult individuals in *C. compressirostris* (EOD duration 0.4 ms in juvenile and 0.4 ms in adult), *C. rhynchophorus* (EOD duration 5 ms in juvenile and 40 ms in adult) and their hybrids (EOD duration 0.4 ms in juvenile and 4 ms in adult). Differentially expressed genes between juveniles and adults were highly enriched in “membrane”, “plasma membrane” and “cytoplasm” Go Ontology terms. We detected several potassium channel-related genes (e.g. *KCNJ2*, *ADCYAP1*) that are potentially involved in the EOD development during ontogeny. The alleles from *C. compressirostris* generally show dominant expression in the hybrid at juvenile and adult life stages. *KCNJ2* is the only gene that exhibits allelic dominance of the *C. rhynchophorus* allele, and has an increasing expression during ontogeny in this allele. This suggests that the EOD development in hybrids could be related to the increasing allelic expression of the *C. rhynchophorus* allele under the scenario of overall dominance of *C. compressirostris* alleles. Our study sheds light in the evolution of the electric organ discharge in electric fishes and on the potential role of introgressive hybridization in complex phenotypic traits.

**Key words:** *Campylomormyrus*, hybrid, EOD, allele specific expression, gene expression, ontogeny.

## INTRODUCTION

Successful hybridization is a rare phenomenon among animals, but hybrids have played an important role in understanding vertebrate evolution (Burke and Arnold 2001), e.g. the emergence of complex phenotypes. Interspecific hybrids result from the mating between closely related species, and contain different ‘parental genomes’ within the same nucleus, referred to as subgenomes. Before the subgenomes become united in the hybrid, each of them has had an

independent evolution (Comai 2005; Schiavinato et al. 2021). The phenotypes in hybrids may range from intermediate forms to transgressive novel traits (Janko et al. 2021). This is occasionally associated with the function of the two subgenomes, in which segments in one subgenome could be selectively reduced or even eliminated causing considerable changes in hybrids (Burke and Arnold 2001; El-mihoub et al. 2006; Janko et al. 2021).

One possible way to understand subgenome evolution is to detect the allele specific expression or the imbalance between the expression levels of the alleles from the parental species in the hybrids (Crowley et al. 2015; Shao et al. 2019). Allele specific expression has been identified in hybrids and polyploids of many animals and plants (Quinn et al. 2014; Crowley et al. 2015; Boutet et al. 2016; Knowles et al. 2017; Cooper and Shaffer 2021; Zhang et al. 2023). In the fish hybrid of *Megalobrama amblycephala* and *Culter alburnus*, the genome-wide transcriptional analysis revealed that the asymmetric expression of alleles could counteract deleterious effects of the subgenomes, and hence improve the adaptability of novel hybrids (Ren et al. 2019).

The African weakly electric fish (Mormyridae) possess a “magic trait” to produce and perceive electric signals (Feulner et al. 2009). Mormyrids use their electric organ to generate electric organ discharge (EOD) for objects sensing and electrocommunication in a social context. The larval electric organ is found in the deep lateral muscle while the adult electric organ is located in the caudal peduncle (Nguyen et al. 2017). The EOD is mainly species-specific and varies in shape and waveform. Among adult *Campylomormyrus*, the short duration EOD represents the ancestral state (plesiomorphic) while the long duration EODs (including medium) are considered as derived (apomorphic) features (Kirschbaum et al. 2016).

The EOD in mormyrids is a compound action potential from the simultaneously firing electrocytes (Mills and Zakon 1991). The depolarization and repolarization phases of an action potential are

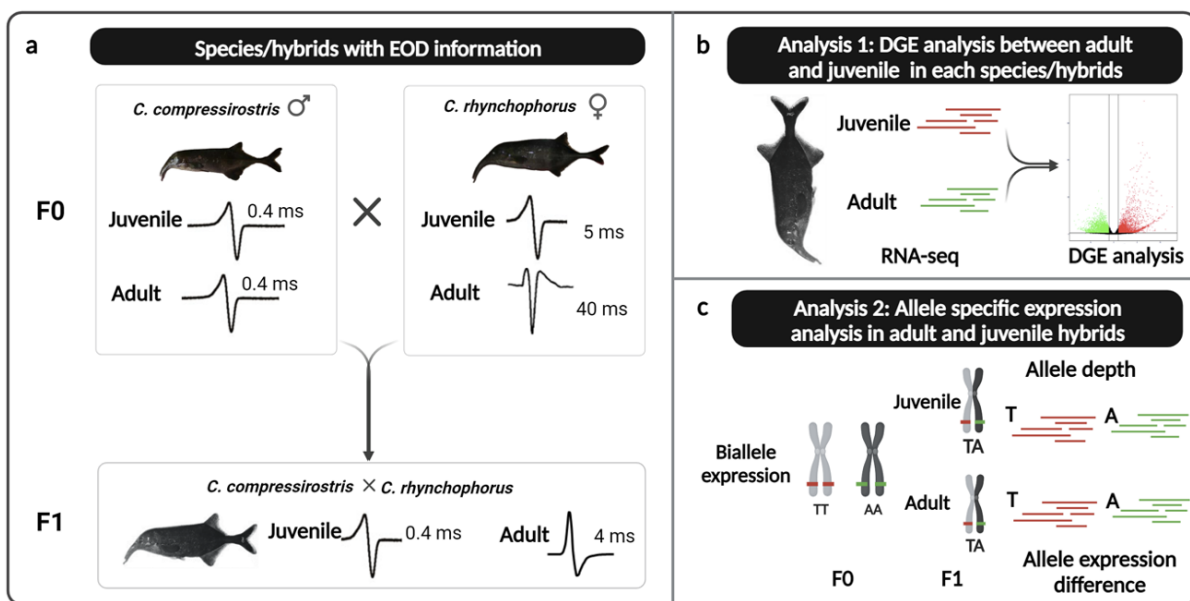
mediated by multiple, interacting, inward and outward ion currents (Stoddard 2008; Jeevaratnam et al. 2018). In *Campylomormyrus*, the diverged EOD waveform might relate to the strength of potassium currents (Cheng et al. 2023a). The inwardly rectifying potassium channel, which is encoded by gene *KCNJ2*, possibly regulates the action potential duration with higher expression in elongated EOD species (Cheng et al. 2023a). In addition, the voltage-gated potassium channel gene *KCNA7a*, which mediates outward potassium currents, also potentially contribute to the EOD duration in weakly electric fish, as has been found in *Gymnarchus* and *Brienomyrus* (Swapna et al. 2018).

EOD variation not only exists in different species, but also in different life stages. Developmental change in electrocytes includes cell growth, formation of papillae (in *C. rhynchophorus*, *C. numenius* and hybrid *C. compressirostris* x *C. rhynchophorus*) and other structural features (Nguyen et al. 2020; Korniienko et al. 2021). Previous investigations on the ontogeny of the EOD in genus *Campylomormyrus* species showed that *C. compressirostris*, *C. tamandua*, *C. tshokwe*, *C. rhynchophorus* share short duration EODs in the early juvenile stage (Kirschbaum et al. 2016; Nguyen et al. 2017; Nguyen et al. 2020; Korniienko et al. 2021). In *C. compressirostris*, the EOD stays consistent during development, while other studied species undergo multiple alterations until they reach the adult EOD. Developmental EOD variation was also found in other mormyrid species, e.g. *Mormyrops* and *Paramormyrops* (Nguyen et al. 2020).

Successful breeding experiments have been performed in *Campylomormyrus* species. They encompass also in interspecific hybrids, although hybrids are rarely be found in the natural habitat (Kirschbaum et al. 2016; Nguyen et al. 2017; Nguyen et al. 2020). Taking advantage of this, we are able to observe the EOD variation in purebred species and hybrids during ontogeny. In most cases, hybrids show an intermediate EOD phenotype compared to their parent species



(Kirschbaum et al. 2016). *Campylomormyrus* hybrids also start with similar short duration juvenile EOD, and reach an adult EOD after several changes in juvenile stages (Kirschbaum et al. 2016). However, the apomorphic feature (medium and long duration EOD) is always dominant over the plesiomorphic (short duration EOD) trait (Kirschbaum et al. 2016). The adult EOD of hybrid *C. compressirostris* x *C. rhynchophorus* is similar to the EOD of 6 cm sized juveniles of *C. rhynchophorus* (Kirschbaum et al. 2016).



**Fig. 1** Electric organ discharge (EOD) of studied *Campylomormyrus* species and hybrids and working flow in this study.

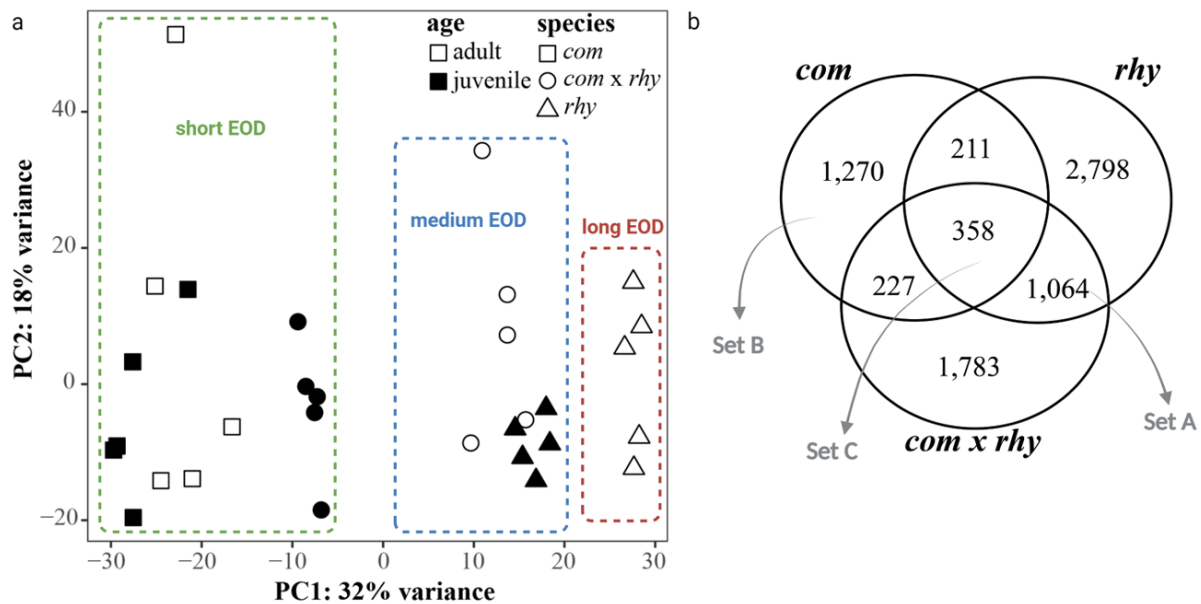
**a** Species/hybrids samples used in the study have EODs with different duration and shape. **b** Differential gene expression (DGE) analysis between juvenile and adult for each species/hybrids. **c** Allelic specific expression analysis in adult and juvenile hybrids. Only the homozygous bialleles in F0 were preserved and the allelic depth was counted for allelic specific expression in two life stages of the hybrids.

This study focuses on the gene expression in juvenile and adult weakly electric fish in both purebred species and hybrids. Our objective is to address the role of hybridization-derived genetic architecture on the expression of complex phenotypes by identifying genes contributing to EOD

development during the ontogeny of parental species, as well as unravelling allele specific expression during hybrid development. We included two life stages samples (adult and ~ 7 cm juvenile) from two F0 species and their F1 hybrid in *Campylomormyrus* (Fig. 1a): *C. compressirostris* (*com*, 0.4 ms short and biphasic adult EOD, 0.4 ms short and biphasic juvenile EOD), *C. rhynchophorus* (*rhy*, 40 ms long and triphasic adult EOD, 5 ms medium and biphasic juvenile EOD), and their hybrid *C. compressirostris* ♂ x *C. rhynchophorus* ♀ (*com* x *rhy*, 4 ms medium and biphasic adult EOD, 0.4 ms short and biphasic juvenile EOD).

## RESULTS

The overall gene expression patterns from both juvenile and adult RNA-seq were analyzed by principal component analysis (PCA, Fig. 2a). This PCA plot preliminary supported a distribution



**Fig. 2** Differential gene expression analysis of the comparison between adult and juvenile species/hybrids.

**a** Principal component analysis (PCA) of gene expression levels in all juvenile and adult species/hybrids.  
**b** Venn Diagram graph for differentially expressed genes (DEGs) between juvenile and adult shared in three species/hybrids. All DEGs have  $|\log_2(\text{fold change})| > 1$  and  $p\text{-value} < 0.05$ .

pattern based on EOD duration in PC1, which explained 35% of variation. The expression pattern from both adult and juvenile *com* samples were clustered together, corresponding with the EOD consistency during the development in this species. In contrast, the adult EODs in the *rhy* and F1 hybrid exhibit shape variation and duration elongation compared with juvenile EODs, and there is a separation between juvenile and adult RNA-seq data of these species in the PCA plot.

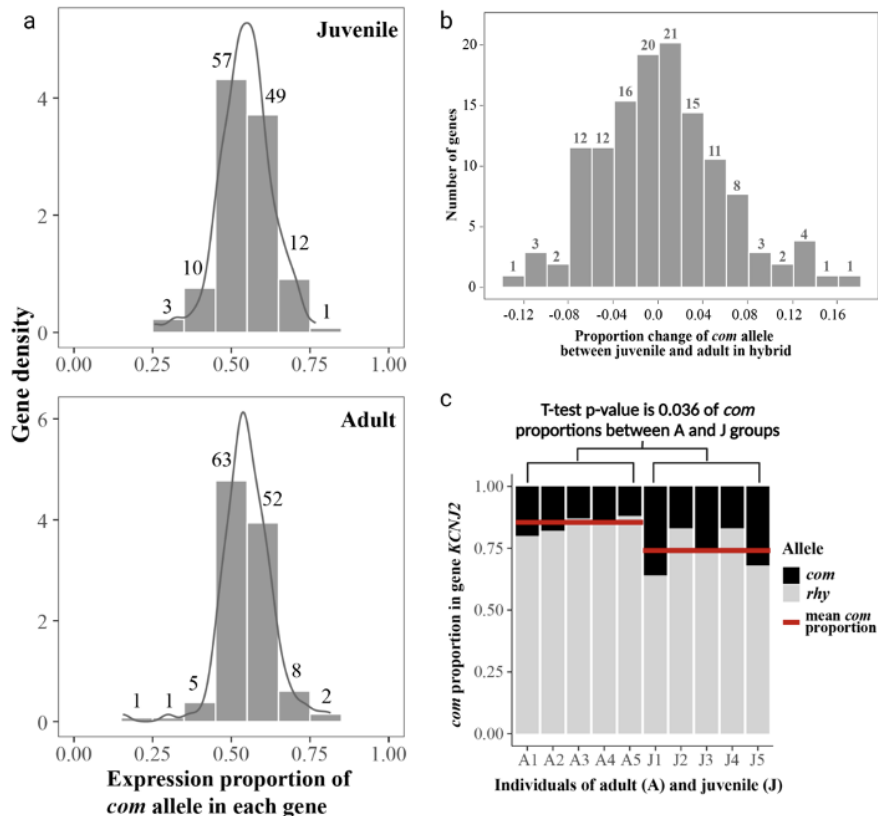
### **Genes relative to electric organ development during ontogeny**

We performed gene expression pairwise comparisons between adult and juvenile electric organ for each species and the hybrids separately (Fig. 1b). We specifically looked at the differentially expressed genes (DEGs) that overlapped among different species/hybrids (Fig. 2b). In *rhy* and the hybrid *com* x *rhy*, the EOD went through shape change and duration elongation from juvenile to adult, while EOD stayed constant in *com*. Therefore, we focused on three sets of DEGs: set A contains the shared DEGs only in *rhy* and the hybrids, representing the DEGs potentially associated to EOD change during ontogeny; set B contains the DEGs uniquely in *com*, representing the DEGs corresponding to an EOD consistency during ontogeny; set C contains the DEGs shared in all species and hybrid, representing the DEGs corresponding to general electric organ development. In total, we identified 1,064 DEGs for set A, 1,270 for set B and 358 from for C (Supplementary Table 1). We blasted all the DEGs against the *nr* database at the National Center for Biotechnology Information (NCBI). Three sets of DEGs were significantly enriched in “membrane”, “plasma membrane”, “cytoplasm”, and “extracellular space/matrix/region” Gene Ontology (GO) terms (Supplementary Table 2, 3, 4).

In set A, the signaling genes *ADCYAP1* showed high overexpression in the adult electric organ. A cytoskeletal and sarcomeric gene *FILIP1* was slightly up-regulated in the adult electric organ. We also found five potassium channel genes (two copies of *KCNA7a*, *KCNH8*, *KCNK1* and *KCNC2*)



i.e., an expression proportion of the *com* allele around 0.45~0.55 in juvenile and adult hybrid, representing 43.2% and 47.7% of total genes, respectively (Fig. 3a). Interestingly, both juvenile and adult hybrids showed a tendency towards overexpression of the *com* allele, i.e., 46.9% of the total genes had a *com* allele expression proportion over 0.55. Only 9.8% of genes showed a bias towards *rhy* allele expression (i.e., *com* allele proportion less than 0.45) in juveniles, and only 5.3% in adult hybrids (Fig. 3a).



**Fig. 3** Proportion of *com* allele expression separate for adults and juveniles, and *com* allele expression change between juvenile and adult in hybrids.

**a** Gene density (y-axis) from juvenile and adult stages in the hybrid. The x-axis shows the expression proportion of the allele stemming from one parental species (*com*). Numbers above the bars represent the number of genes in the respective proportion ranges. **b** *Com* allele proportion change from juvenile to adult stage, the number of genes is showed above every bar in the respective proportion variation range. **c** Allelic proportion of *KCNJ2* gene in each individual of juvenile and adult hybrids. The t-test p-value for *com* allele proportion between juvenile and adult individuals indicates a significant difference among ontogenetic stages, i.e., a larger expression bias towards the *rhy* allele in adults.

From those 132 genes, 50% showed increasing *com* allele expression proportion throughout ontogeny. Only 12 had the net proportion change between juveniles and adults over 10% (8 increased, 4 decreased expression bias in adults; Fig. 3b, Table 2).

**Table 2** 12 genes show a *com* proportion difference over 10% between juvenile and adult hybrid.

Gene ID in annotation	Gene	Highlights of Predicted Function	Gene Description	Category	<i>com</i> proportion in hybrid		<i>com</i> proportion variation
					juvenile	adult	
maker-ptg002464l-snap-gene-0.36-mRNA-1	<i>CEP68</i>	enables protein domain specific binding and kinase binding activity	centrosomal protein 68	other	0.40	0.57	0.17
maker-ptg000399l-augustus-gene-13.63-mRNA-1	<i>HIP1</i>	membrane-associated protein	huntingtin-interacting protein 1-related protein	other	0.44	0.59	0.15
maker-ptg000051l-snap-gene-130.10-mRNA-1	<i>TULP4</i>	protein ubiquitination	tubby-related protein 4	other	0.40	0.53	0.13
maker-ptg000881l-augustus-gene-0.121-mRNA-1	<i>FILIP1</i>	acting through a filamin-A/F-actin axis	filamin-A-interacting protein 1	other	0.48	0.61	0.13
maker-ptg000312l-snap-gene-7.122-mRNA-1	<i>PLEKHA8</i>	cargo transport protein for apical transport from Golgi complex	pleckstrin homology domain containing A8	other	0.60	0.47	-0.13
maker-ptg000160l-snap-gene-11.44-mRNA-1	<i>CHRND</i>	subunit of the nicotinic acetylcholine receptor ion channel	acetylcholine receptor subunit delta	transmembrane ion transport	0.69	0.82	0.13
maker-ptg001623l-snap-gene-3.166-mRNA-1	<i>ARHGEF26</i>	activates RhoG GTPase by promoting the exchange of GDP by GTP	Rho guanine nucleotide exchange factor 26	other	0.33	0.46	0.13
maker-ptg001894l-augustus-gene-1.273-mRNA-1	<i>HEPH1</i>	transmembrane transport of various molecules	pannexin 1	transmembrane ion transport	0.62	0.74	0.12
maker-ptg001572l-augustus-gene-4.62-mRNA-1	<i>NFU1</i>	iron-sulfur cluster biogenesis	NFU1 iron-sulfur cluster scaffold homolog, mitochondrial-like	other	0.56	0.68	0.12
maker-ptg000959l-snap-gene-7.19-mRNA-1	<i>PLPP3</i>	converts phosphatidic acid to diacylglycerol	phospholipid phosphatase 3	other	0.66	0.55	-0.12
maker-ptg000608l-snap-gene-5.61-mRNA-1	<i>SYNE2</i>	nuclear outer membrane protein that binds cytoplasmic F-actin	spectrin repeat containing nuclear envelope protein 2	other	0.61	0.51	-0.10
maker-ptg000265l-est_gff_est2genome-gene-6.33-mRNA-1	<i>KCNJ2</i>	allow potassium to flow into a cell rather than out of a cell, probably participates in establishing action potential waveform	inward rectifier potassium channel 2	transmembrane ion transport	0.26	0.16	-0.10

Those genes with a lower *com* allele proportion in the adult hybrid were *PLEKHA8*, *SYNE2*, *PLPP3* and *KCNJ2*. We specifically looked at the allelic expression in each individual at gene *KCNJ2* since this gene has been found to be up-regulated in elongated EOD duration *Campylomormyrus* species (Cheng et al. 2023a) and also in adults *rhy* and hybrids *com* x *rhy* compared with their juveniles (DEG in set A). In this study, the *com* allele of *KCNJ2* gene showed higher expression proportion (0.26) in juvenile hybrids, compared to adults (0.16; Table 2; t-test of arcsine transformed *com* proportion p-value=0.036; Fig. 3c). The *rhy* allele in *KCNJ2* gene is more dominant in the adult electric organ, indicating a considerable allelic expression shift during

hybrid ontogeny. *Com* allele in both genes *CHRND* and *FILIP1* showed increasing expression during ontogeny (Table 2), and the expression of the former was also increased in the skeletal muscle of adult hybrids (Cheng et al. 2023a). Interestingly, the expression of *CHRND* gene was up-regulated in the electric organ during the development of *com* species (set B), while *FILIP1* was up-regulated in both *rhy* and hybrid (set A).

## DISCUSSION

### Allele specific expression reveals subgenome effect in the hybrids

The weakly electric mormyrids usually possess a larval electric organ in early age and an adult electric organ that persists in adults. Adult EODs are generally different from the larval and juvenile EODs during ontogeny. For example, the species *rhy* possesses a relatively short biphasic juvenile EOD (0.5 ms), and gradually reaches an adult long triphasic EOD (40 ms) during development (Nguyen et al. 2020). Deviating from this, the EOD in species *com* shows consistency during ontogeny, remaining very short (0.4 ms) and biphasic (Kirschbaum et al. 2016). In the hybrid *com* x *rhy*, the early juvenile EOD was quite similar in shape and duration to the *com* EOD. In adulthood, however, the hybrid EOD was still biphasic but elongated (5 ms), resembling the EOD of a *rhy* juvenile of 6~7 cm body length (Kirschbaum et al. 2016; Nguyen et al. 2020). It suggested that the EOD development in this hybrid is more close to the species *rhy* with an EOD considered apomorphic, instead of the plesiomorphic character state of species *com* (Kirschbaum et al. 2016; Cheng et al. 2023a).

Despite the hybrid showing phenotypic changes during ontogeny in the EOD similar to *rhy*, at the transcriptional level there is a general expression bias towards *com* in the hybrid at both life stages (Fig. 3a). The higher *com* allele expression also occurred in the adult skeletal muscle of hybrid

*com* x *rhy* as well as in the adult skeletal muscle and electric organ of hybrid *C. compressirostris* x *C. tshokwe* (Cheng et al. 2023a). The electric organ of both adult hybrids showed an even higher expression bias towards the *com* allele (Cheng et al. 2023a). In addition, eight genes had an increasing expression bias towards the *com* allele throughout ontogeny with an increase in bias over 10% in adult hybrid (eight genes) compared to the juvenile, whereas four genes showed a decrease in expression bias throughout ontogeny (Table 2). This suggests that the subgenome from *com* species might be more dominant in the hybrid *com* x *rhy* in general (especially in the electric organ), while the EOD phenotype resembles the one of the *rhy* species.

Interestingly, *KCNJ2* was the only gene that showed significant expression dominance of the *rhy* allele in the hybrid *com* x *rhy* at both juvenile and adult stages, and the dominance was increasing throughout ontogeny (Fig. 3c). The inwardly rectifying potassium channel gene (*KCNJ2*) is considered as a powerful candidate gene for regulating the EOD duration in *Campylomomyrus* since its expression is up-regulated in species with elongated EOD (Cheng et al. 2023a). The expression of this gene did also increase during ontogeny in *rhy* and in the hybrids (set A; Table 1) which both exhibit an increasing EOD duration elongation throughout ontogeny. The inward potassium channel encoded by *KCNJ2* gene can pass positive charged current more easily in the inward than the outward direction (Hibino et al. 2010). This channel is considered as important in regulating neuronal activity by stabilizing the resting membrane potential. Moreover, it may contribute to the shaping of the initial depolarization and the final repolarization step during the action potential, as evidenced in human cardiomyocytes (Dhamoon and Jalife 2005; Jeevaratnam et al. 2018).

The up-regulated *KCNJ2* in adult *rhy*/hybrids and increased dominance of the *rhy* allele in adult further supported the potential function of *KCNJ2* gene in EOD duration regulation. In addition,



even though the subgenome from *com* species was more dominant in the hybrid (especially in the electric organ), the EOD phenotype (i.e. EOD duration) appears affected by the *rhy* allele in *KCNJ2* gene during ontogeny.

Juvenile individuals (J2 and J4 in Fig. 3c) showed *rhy* allele expression proportion similar to the adult. The EOD duration in J2 already showed slight elongation (around 0.45 ms compared to 0.4 ms). It is likely that the expression bias towards the *rhy* allele occurs, when the juvenile EODs starts to elongate, and is maintained until adulthood.

The gene *CHRND* encodes for a cholinergic receptor nicotinic delta subunit. It binds acetylcholine and affects all subunits to open an ion-conducting channel across the plasma membrane (Shen et al. 2016). This gene was not only up-regulated in the adult *com* species, relative to the juvenile, but also showed an increasing expression bias towards the *com* allele in both electric organ and skeletal muscle in the hybrid during ontogeny. This pattern points towards a cis-regulation of this gene in *Campylomormyrus* species and their hybrids. Unfortunately, the function of this gene in electric fish is still unknown.

### **Electric organ discharge development during ontogeny in *Campylomormyrus***

Another goal of this study was to identify the candidate genes involved in the EOD alteration during ontogeny (in *rhy* and hybrids), relative to the EOD consistency in *com*. The depolarization and repolarization of EOD is associated with sodium and potassium currents across the plasma membrane. In electric fish, the initial depolarization phase is mainly driven by inward sodium current by voltage-gated sodium channels, and repolarization, ultimately returning the membrane to the resting potential, is principally affected by outward current through voltage-gated potassium channels (Stoddard 2008). Therefore, the potassium channel activity is the essential determinant

of the action potential duration as it limits the repolarization duration (Stoddard 2008; Jeevaratnam et al. 2018). The gene *KCNJ2* was already identified as a powerful candidate gene in regulating the EOD duration during ontogeny (Cheng et al. 2023a). In addition, other genes possibly contribute to the change in EOD duration during ontogenesis.

During the ontogeny of *rhy* and hybrid (set A), we observed several down-regulated voltage-gated potassium channel genes (two copies of *KCNA7a*, *KCNH8*, *KCNCL1* and *KCNC2*) which could possibly reduce the outward potassium current during repolarization. The gene *ADCYAP1* encodes pituitary adenylate cyclase activating polypeptide (PACAP). It has the ability to bind a G protein-coupled receptor PAC1 with relatively high affinity (Johnson et al. 2019). The PACAP/PAC1 receptor signaling is potentially coordinating the function of several ionic channels to regulate neuronal excitability (Hammack et al. 2015), e.g. it has been shown to rapidly inactivate the potassium current in humans. Therefore, the up-regulated *ADCYAP1* might also inactivate the potassium current in electric fish. *KCNS3* forms functional heterotetrameric channels with *KCNBI* and regulates the potassium current activation and deactivation rates of *KCNBI*, as shown in human lens epithelium (Shepard and Rae 1999). Different types of potassium channel may contribute to the EOD shape and duration alteration during ontogeny.

*Com* is the only *Campylomormyrus* species that shows EOD consistence during ontogeny in *Campylomormyrus*. During the ontogeny, the electrocytes undergo cell growth and other structural changes (Nguyen et al. 2020). The membrane capacity may hence be enlarged in following with the growth of electrocytes. It is possible that the DEGs with ontogenetic changes in expression only in *com* (i.e., genes in set B) compensate for this developmental change to maintain a constant EOD in different life stages. The gene *KCNVI* down-regulates the delayed outward rectifier voltage-gated potassium channels, e.g. *KCNB2*, by the formation of heteropolymeric channels

(Hugnot et al. 1996). Differently expressed potassium channel genes (*KCNH1*, *KCNQ4*, *KCNA6a*) might have compensate for the developmental change on electrocytes and consequently stabilized the EOD shape and duration in *com* during ontogeny.

These results support the notion that the equipment of electrocytes in different species/hybrid cohorts and/or at a different developmental stage with a unique set of potassium channels is critical for the duration (and possibly also the shape) of the EOD. The exact function of these potassium channels in this scenario must remain open yet, since detailed information of the electrophysiological properties of the various channel types in these species, e.g. the effects of a different subunit composition for a distinct channel type, are missing. Further studies about the histology and physiology of electric organ development during ontogeny would be essential to understand the mechanisms of EOD development.

## CONCLUSION

An important driver of phenotypic evolution in closely related species is the expression divergence in the genes underlying the respective phenotypic trait. Hybrids possess intermediate or even novel phenotypes because of their specific pattern of gene expression of two subgenomes from their parental species. In our example, the hybrid phenotype (here EOD) in juveniles is similar to *C. compressirostris* and in adults close to *C. rhynchophorus*.

The allele specific expression patterns inferred from hybrid transcriptome data of juvenile and adult electric organs revealed a distinct allelic expression dominance of the alleles from *C. compressirostris* in the hybrid. But our focus phenotype (the EOD) in the hybrids *C. compressirostris* x *C. rhynchophorus* rather resembled the other parental species, i.e., *C. rhynchophorus*. This points towards strong impact of single genes. Indeed, a strong candidate for

having profound influence on the elongated EOD is *KCNJ2*. *KCNJ2* was the only gene that showed significant allelic expression dominance of the *C. rhynchophorus* allele in both juvenile and adult life stages, and the *rhy* allele became increasingly dominant throughout ontogeny. This gene was previously assumed to be involved in the EOD duration regulation in *Campylomormyrus* (Cheng et al. 2023a). In addition, its expression level also increases throughout the ontogeny of both *C. rhynchophorus* and the hybrids, associated with an EOD duration prolongation during ontogeny. Therefore, we hypothesized it as a crucial gene in regulating the EOD duration during the ontogeny in *Campylomormyrus*. Moreover, it likely affects the EOD in the hybrid development under a general expression dominance of the *C. compressirostris* subgenome.

A functional test of the impact of the *KCNJ2* gene in electric fish would be instrumental for a mechanistic understanding of the EOD divergence and development. In addition, a potential backcross experiment of F1 hybrid with *C. compressirostris* species might be interesting to explore if the allelic expression will affect the next generation.

Differentially expressed potassium channel genes, e.g. *ADCYAP1*, *KCNA7a*, *KCNS3*, *KCNC1* are further promising candidates to impact the EOD development or consistency throughout ontogeny.

## **Materials and Methods**

### **Samples and RNA sequencing**

We sampled the electric organ from two F0 species (*C. compressirostris*, *C. rhynchophorus*) and a F1 hybrid (*C. compressirostris* ♂ x *C. rhynchophorus* ♀) at juvenile (around 6 to 7 cm body size) life stage. The juvenile species/hybrid were artificially bred and raised at the University of Potsdam. We included five individuals from each species/the hybrids for mRNA sequencing. We anesthetized the fish specimens using a lethal dose of clove oil, and sampled the electric organs

within 99% ethanol above ice. The extracted electric organ samples were flash frozen in liquid nitrogen and preserved in -80°C.

We applied QIAGEN RNeasy Fibrous Tissue Kit on all juvenile electric organ samples for RNA extraction and further inspected the quantity and quality using a NanoDrop 1000 spectrophotometer (ThermoFischer Scientific, Germany) and an Agilent Bioanalyzer 2100 (Agilent Technologies, USA), respectively. We performed mRNA enrichment via poly (A) capture from the isolated RNA using NEXTflex Poly (A) Beads, and then established the strand-specific transcriptomic libraries through NEXTflex Rapid Directional RNA-Seq Kit (Bioo Scientific, USA).

All constructed libraries were sent to a commercial company (Novogene) for sequencing at 150 bp paired-end reads using an Illumina HiSeq 4000 sequencing system. Raw reads were deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus.

We also included the mRNA sequences of electric organs of adult *C. compressirostris*, *C. rhynchophorus* and their hybrid *C. compressirostris* ♂ x *C. rhynchophorus* ♀ from Cheng et. al (2023a; accessions number: GSE240783). In order to remove the adapter sequences and low quality reads, all sequences were trimmed using a 4 bp sliding window with a mean quality threshold of 25, and a minimum read length of 36 bp by Trimmomatic v0.39 (Bolger et al. 2014). Reads quality was measured by FastQC v0.11.9 (Andrew 2010).

### **Differential gene expression analysis**

We mapped the quality-filtered RNA reads from each samples to the *C. compressirostris* genome (Cheng et al. 2023b) by RSEM (Li and Dewey 2011). The estimated gene level quantification counts were used as input in R/Bioconductor by tximport package. We removed low count ( $\leq 10$ )

and low frequency (less than two replicates) genes to get cleaned reads. Principle component analysis (PCA) was applied for the cleaned and log-transformed counting matrices.

For the purpose of identifying genes that are differentially expressed between adult and juvenile samples, we applied normalized counting matrices to DESeq2 (Love et al. 2014) between adult and juvenile species/hybrids respectively. Differentially expressed genes (DEGs) were identified using  $|\log_2 \text{Folder change (log}_2\text{FC)}| > 1$  and  $p\text{-value} < 0.05$ . A Venn diagram (Hanbo Chen 2011) was compiled to visualize different groups of DEGs across adults and juveniles of both purebred species and their hybrids.

We focused on three DEGs sets (A, B and C). Three sets of DEGs were blasted against the *nr* database from NCBI using blastx with an e-value cutoff  $1e^{-10}$ . Those DEGs were also applied for Gene Ontology (GO) enrichment analysis respectively (Dennis et al. 2003).

### **Allelic specific expression analysis between adult and juvenile hybrids**

Knowing hybrid's allelic expression bias in different life stages can contribute to the understanding of hybrid development. In this study, we focused on genes with SNPs that had fixed bialleles (homozygous) in parental F0 species.

We mapped the cleaned mRNA reads from each sample to the *C. compressirostris* genome by STAR v2.7.7 (Dobin and Gingeras 2015) and sorted the generated bam files from STAR according to the coordinates using SAMtools v1.15 (Danecek et al. 2021). All the sorted bam files were imported to BCFtools v1.9 (Danecek et al. 2021) for variant calling and then filtered following Cheng et al. (2023a). Eventually we acquired high quality bialleles in both adult and juvenile for genes where the parental species are homozygous for different alleles.

We calculated the expression proportion of the allele from *C. compressirostris* (*com*) by assessing the SNPs status in hybrid samples. The average *com* allele proportion was counted to gene level for genes with more than one SNP. The proportion difference between juvenile and adult in hybrid were calculated.

For the gene *KCNJ2*, we counted the proportion from *com* allele in the hybrid individuals. The arcsine-transformed proportions between juveniles and adults were analyzed using a T-test (Kim 2015).

### **Acknowledges**

We thank Dr. Linh Nguyen for fish and hybrid breeding and raising. The project is funded by University of Potsdam.

### **Author contributions**

RT conceived and supervised this study, and provided financial support. FC performed lab work, analysis and drafted the manuscript with the input from DAB, BO, KF and DM.

### **Data availability**

Sequence data have been deposited at NCBI Gene Expression Omnibus under accession GSE240784.

### **REFERENCE**

- Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics*.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

- Boutet G, Alves Carvalho S, Falque M, Peterlongo P, Lhuillier E, Bouchez O, Lavaud C, Pilet-Nayel ML, Rivière N, Baranger A. 2016. SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics* 17.
- Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. *Annu Rev Genet* 35:31–52.
- Cheng F, Dennis AB, Baumann O, Kirschbaum F, Abdelilah-Seyfried S, Tiedemann R. 2023a. Gene and allele specific expression underlying the electric signal divergence in African weakly electric fish. *Unpublished manuscript*.
- Cheng F, Dennis AB, Osuoha J, Canitz J, Kirschbaum F, Tiedemann R. 2023b. A new genome of an African weakly electric fish (*Campylomormyrus compressirostris*, Mormyridae) indicates rapid gene family evolution in Osteoglossomorpha. *BMC Genomics*. 24:129.
- Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet* 6:836–846.
- Cooper RD, Shaffer HB. 2021. Allele-specific expression and gene regulation help explain transgressive thermal tolerance in non-native hybrids of the endangered California tiger salamander (*Ambystoma californiense*). *Mol Ecol* 30:987–1004.
- Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, Wang JR, Morgan AP, Calaway JD, Aylor DL, et al. 2015. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat Genet*. 47:353–360.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:1–4.
- Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4.
- Dhamoon AS, Jalife J. 2005. The inward rectifier current (IK1) controls cardiac excitability and is involved in arrhythmogenesis. *Heart Rhythm* 2:316–324.
- Dobin A, Gingeras TR. 2015. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* 51:11.14.1-11.14.19.
- El-mihoub TA, Hoppgood AA, Nolle L, Battersby A. 2006. Hybrid Genetic Algorithms : A Review. *Engineering Letters* 13:124–137.
- Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R. 2009. Magic trait electric organ discharge (EOD): Dual function of electric signals promotes speciation in African weakly electric fish. *Commun. Integr. Biol.* 2:329–331.
- Hammack SE, May V, Hall JD. 2015. Pituitary adenylate cyclase activating polypeptide in stress- related disorders: data convergence from animal and human studies. *Biol Psychiatry* 78:167–177.
- Hanbo Chen PCB. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* 12:1–7.
- Hibino H, Inanobe A, Furutani K, Murakami S, Findlay I, Kurachi Y. 2010. Inwardly rectifying potassium channels: Their structure, function, and physiological roles. *Physiol Rev* 90:291–366.



- Hugnot JP, Salinas M, Lesage F, Guillemare E, De Weille J, Heurteaux C, Mattéi MG, Lazdunski M. 1996. Kv8.1, a new neuronal potassium channel subunit with specific inhibitory properties towards Shab and Shaw channels. *EMBO Journal* 15:3322–3331.
- Janko K, Bartoš O, Kočí J, Roslein J, Drdová EJ, Kotusz J, Eisner J, Mokrejš M, Štefková-Kašparová E. 2021. Genome Fractionation and Loss of Heterozygosity in Hybrids and Polyploids : Mechanisms , Consequences for Selection , and Link to Gene Function. *Mol Biol Evol* 38:5255–5274.
- Jeevaratnam K, Chadda KR, Huang CLH, Camm AJ. 2018. Cardiac Potassium Channels: Physiological Insights for Targeted Therapy. *J Cardiovasc Pharmacol Ther* 23:119–129.
- Johnson GC, May V, Parsons RL, Hammack SE. 2019. Parallel signaling pathways of pituitary adenylate cyclase activating polypeptide (PACAP) regulate several intrinsic ion channels. *Ann N Y Acad Sci*. 1455:105–112.
- Kim TK. 2015. T test as a parametric statistic. *Korean J Anesthesiol* 68:540–546.
- Kirschbaum F, Nguyen L, Baumgartner S, Chi HWL, Wolfart R, Elarbani K, Eppenstein H, Korniienko Y, Guido-Böhm L, Mamonekene V, et al. 2016. Intra-genus (*Campylomormyrus*) and intergenus hybrids in mormyrid fish: Physiological and histological investigations of the electric organ ontogeny. *J Physiol Paris* 110:281–301.
- Knowles DA, Davis JR, Edgington H, Raj A, Favé MJ, Zhu X, Potash JB, Weissman MM, Shi J, Levinson DF, et al. 2017. Allele-specific expression reveals interactions between genetic variation and environment. *Nat Methods*. 14:699–702.
- Korniienko Y, Tiedemann R, Vater M, Kirschbaum F. 2021. Ontogeny of the electric organ discharge and of the papillae of the electrocytes in the weakly electric fish *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *Journal of Comparative Neurology* 529:1052–1065.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.*:1–16.
- Love MI, Anders S, Huber W. 2014. Differential analysis of count data - the DESeq2 package.
- Mills A, Zakon HH. 1991. Chronic androgen treatment increases action potential duration in the electric organ of *Sternopygus*. *Journal of Neuroscience* 11:2349–2361.
- Nguyen L, Mamonekene V, Vater M, Bartsch P, Tiedemann R, Kirschbaum F. 2020. Ontogeny of electric organ and electric organ discharge in *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *Journal of Comparative Physiology A* 206:453–466.
- Nguyen L, Paul C, Mamonekene V, Bartsch P, Tiedemann R, Kirschbaum F. 2017. Reproduction and development in some species of the weakly electric genus *Campylomormyrus* (Mormyridae, Teleostei). *Environ Biol Fishes* 100:49–68.
- Stoddard P K, Markham M R. 2008. Signal Cloaking by Electric Fish. *Bioscience*. 58:415–425.
- Quinn A, Juneja P, Jiggins FM. 2014. Estimates of allele-specific expression in *Drosophila* with a single genome sequence and RNA-seq data. *Bioinformatics* 30:2603–2610.

- Ren L, Li W, Qin Q, Dai H, Han F, Xiao J, Gao X, Cui J, Wu C, Yan X, et al. 2019. The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* × *Culter alburnus*. *Genome Res* 29:1805–1815.
- Schiavinato M, Bodrug-Schepers A, Dohm JC, Himmelbauer H. 2021. Subgenome evolution in allotetraploid plants. *Plant Journal* 106:672–688.
- Shao L, Xing F, Xu C, Zhang Qinghua, Che J, Wang X, Song J, Li X, Xiao J, Chen LL, et al. 2019. Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc Natl Acad Sci U S A* 116:5653–5658.
- Shen X-M, Okuno T, Milone M, Otsuka K, Takahashi K, Komaki H, Giles E, Ohno K, Engel AG. 2016. Mutations causing slow-channel myasthenia reveal that a valine ring in the channel pore of muscle AChR is optimized for stabilizing channel gating. *Hum Mutat.* 37:1051–1059.
- Shepard AR, Rae JL. 1999. Electrically silent potassium channel subunits from human lens epithelium. *Am J Physiol Cell Physiol* 277.
- Swapna I, Ghezzi A, York JM, Markham MR, Halling DB, Lu Y, Gallant JR, Zakon HH. 2018. Electrostatic Tuning of a Potassium Channel in Electric Fish. *Current Biology* 28:2094–2102.
- Zhang Q, Ye Z, Wang Y, Zhang X, Kong W. 2023. Haplotype-Resolution Transcriptome Analysis Reveals Important Responsive Gene Modules and Allele-Specific Expression Contributions under Continuous Salt and Drought in *Camellia sinensis*. *Genes (Basel)* 14:1417.

## Supplementary Table 1

Supplementary Table 1 Differentially expressed candidate genes with in set A, B and C

Gene ID in annotation	Gene	Highlights of Predicted Function	Gene Description	Category	log2FC			
					com	rhy	Set com x rhy	
maker-ptg0000851-snap-gene-19.30-mRNA-1	MYH2	myosin heavy chain	myosin heavy chain, fast skeletal muscle	cytoskeletal & sarcomeric	-	4.05	5.22	A
maker-ptg000572l-augustus-gene-14.21-mRNA-1	SMYHC1	myosin heavy chain beta isoform	myosin-7	cytoskeletal & sarcomeric	-	2.97	3.28	A
maker-ptg000734l-snap-gene-4.160-mRNA-1	MYO5C	unconventional myosin-Vc	myosin VC	cytoskeletal & sarcomeric	-	2.80	2.49	A
maker-ptg001003l-snap-gene-2.16-mRNA-1	MYO1E	unconventional myosin; actin-based motor protein	unconventional myosin-le	cytoskeletal & sarcomeric	-	2.53	1.60	A
maker-ptg000881l-augustus-gene-0.121-mRNA-1	FILIP1	acting through a filamin-A/F-actin axis	filamin-A-interacting protein 1	cytoskeletal & sarcomeric	-	1.66	1.06	A
maker-ptg000548l-snap-gene-0.90-mRNA-1	PNKD	probable hydrolase; activation of the NF-kappa-B signaling pathway	paroxysmal nonkinetogenic dyskinesia	other	-	1.70	1.30	A
snap_masked-ptg000585l-processed-gene-1.0-mRNA-1	HS3ST1	a member of the heparan sulfate biosynthetic enzyme family	heparan sulfate glucosaminase 3-O-sulfotransferase 1	other	-	7.01	6.40	A
maker-ptg000174l-snap-gene-2.8-mRNA-1	H2-Aa	antigen processing and activation	H-2 class II histocompatibility antigen, A-U alpha chain	other	-	5.44	6.05	A
maker-ptg000827l-snap-gene-1.93-mRNA-1	OPRM1	G protein-coupled receptor for natural and synthetic opioids	mu-type opioid receptor	other	-	6.39	5.06	A
maker-ptg000528l-augustus-gene-4.129-mRNA-1	CCDC105	located in extracellular exosome	coiled-coil domain containing 105	other	-	5.26	5.23	A
snap_masked-ptg000142l-processed-gene-0.216-mRNA-1	CDH17	calcium-dependent cell adhesion protein	cadherin-17	other	-	1.38	2.84	A
maker-ptg002338l-snap-gene-0.93-mRNA-1	KCTD14	protein-containing complex assembly	potassium channel tetramerization domain containing 14	other	-	1.38	1.75	A
maker-ptg002895l-snap-gene-2.33-mRNA-1	ICAM1	binding of a cell to another cell or to the extracellular matrix	intercellular adhesion molecule 1	other	-	-6.12	-5.82	A
maker-ptg000974l-snap-gene-8.9-mRNA-1	FAM20C	post-translational protein modification	extracellular serine/threonine protein kinase FAM20C	other	-	5.30	5.62	A
maker-ptg000130l-est_gff_est2genome-gene-5.10-mRNA-1	CLDN7	integral membrane protein; tight junction component	claudin-7	other	-	5.11	5.62	A
maker-ptg000061l-augustus-gene-13.21-mRNA-1	CDH26	calcium-dependent cell adhesion protein	cadherin-26	other	-	3.59	1.29	A
snap_masked-ptg000633l-processed-gene-29.4-mRNA-1	CADPS2	positive regulation of exocytosis	calcium dependent secretion activator 2	other	-	1.95	2.35	A

maker-ptg0004051-snap-gene-38.13-mRNA-1	<i>SMOC2</i>	extracellular matrix organization	SPARC related modular calcium binding 2	other	-	1.45	1.11	A
maker-ptg0002221-snap-gene-9.11-mRNA-1	<i>CDH20</i>	calcium-dependent cell adhesion protein	cadherin-20	other	-	-4.14	-4.01	A
maker-ptg001524l-augustus-gene-3.4-mRNA-1	<i>TMEM47</i>	regulates cell junction organization	transmembrane protein 47	other	-	3.10	4.40	A
maker-ptg000852l-augustus-gene-40.4-mRNA-1	<i>MARVELD2</i>	cell-cell junction organization	MARVEL domain-containing protein 2	other	-	1.29	6.05	A
maker-ptg000032l-augustus-gene-10.24-mRNA-1	<i>CAB39</i>	protein kinase activator activity	calcium-binding prote in 39	signaling	-	1.45	1.21	A
maker-ptg000974l-augustus-gene-10.57-mRNA-1	<i>PVALB2</i>	high affinity calcium ion-binding protein	parvalbumin-2-like	signaling	-	-5.02	-8.72	A
snap_masked-ptg001873l-processed-gene-0.7-mRNA-1	<i>NXP1</i>	signaling molecules that resemble neuropeptides	neurexophilin-1	signaling	-	6.11	5.51	A
maker-ptg000267l-snap-gene-2.92-mRNA-1	<i>PIEZO2</i>	mechanosensitive ion channel	piezo type mechanosensitive ion channel component 2	signaling	-	1.16	1.34	A
maker-ptg000084l-augustus-gene-28.11-mRNA-1	<i>CAMK1D</i>	Ca2+/calmodulin-dependent protein kinase activity	calcium/calmodulin dependent protein kinase ID	signaling	-	-1.24	-1.41	A
maker-ptg001479l-augustus-gene-0.6-mRNA-1	<i>ADCYA1</i>	activation of adenylate cyclase activity	adenylate cyclase activating polypeptide 1	signaling	-	8.95	5.56	A
maker-ptg000335l-snap-gene-16.83-mRNA-1	<i>KCNM3</i>	modulates the delayed rectifier voltage-gated potassium channel activation and deactivation rates of KCNB1	potassium voltage-gated channel subfamily 5 member 3	transmembrane ion transport	-	4.07	6.07	A
maker-ptg000878l-augustus-gene-3.11-mRNA-1	<i>KCNQ5a</i>	form M-type outward potassium current when depolarization, counteracts sodium influx to prevent action potential	potassium voltage-gated channel subfamily KQT member 5	transmembrane ion transport	-	5.35	1.87	A
maker-ptg000265l-est_gff_est2genome-gene-6.33-mRNA-1	<i>KCNJ2</i>	allow potassium to flow into a cell; probably participates in establishing action potential waveform	inward rectifier potassium channel 2	transmembrane ion transport	-	1.13	1.73	A
maker-ptg000028l-snap-gene-81.8-mRNA-1	<i>KCNA7a_2</i>	voltage-dependent potassium ion permeability of excitable membranes	potassium voltage-gated channel subfamily A member 7a	transmembrane ion transport	-	-2.10	-1.11	A
maker-ptg000028l-snap-gene-81.10-mRNA-1	<i>KCNA7a_1</i>	voltage-dependent potassium ion permeability of excitable membranes	potassium voltage-gated channel subfamily A member 7a	transmembrane ion transport	-	-1.06	-2.38	A
maker-ptg000922l-snap-gene-13.16-mRNA-1	<i>KCNH8</i>	elicits a slowly activating, outward rectifying current	potassium voltage-gated channel subfamily H member 8	transmembrane ion transport	-	-2.36	-1.39	A
maker-ptg001310l-augustus-gene-10.4-mRNA-1	<i>KCNK2</i>	regulation of the fast action potential repolarization	potassium voltage-gated channel subfamily C member 2	transmembrane ion transport	-	-1.04	-1.41	A
maker-ptg000898l-augustus-gene-5.59-mRNA-1	<i>KCNK1</i>	regulation of the fast action potential repolarization	potassium voltage-gated channel subfamily C member 1	transmembrane ion transport	-	-3.13	-3.18	A
maker-ptg000078l-augustus-gene-7.51-mRNA-1	<i>KCNIP1</i>	inactivating A-type potassium channels	Kv channel-interacting protein 1	transmembrane ion transport	-	-4.23	-2.20	A
maker-ptg0000897l-augustus-gene-1.69-mRNA-1	<i>SLC6a20</i>	amino acid transmembrane transporter	solute carrier family 6 member 20	transmembrane ion transport	-	3.19	3.04	A

maker-ptg000611l-snap-gene-3.83-mRNA-1	<i>SLC7a7</i>	amino acid transmembrane transporter	solute carrier family 7 member 7	-	2.32	2.78	A
maker-ptg000326l-snap-gene-3.64-mRNA-1	<i>SLC13a1</i>	sodium:sulfate symporter	solute carrier family 13 member 1	-	2.14	2.15	A
maker-ptg000802l-augustus-gene-1.88-mRNA-1	<i>SLC03a1</i>	sodium-independent organic anion transmembrane transporter	solute carrier organic anion transporter family member 3A1	-	1.89	1.92	A
snap_masked-ptg001837l-processed-gene-1.59-mRNA-1	<i>SLC25a23</i>	adenyl nucleotide antiporter in the inner mitochondrial membrane	solute carrier family 25 member 23	-	-1.37	-2.20	A
maker-ptg000361l-snap-gene-18.27-mRNA-1	<i>SLC22a23</i>	antiporters to transport organic ions across cell membranes	solute carrier family 22 member 23	-	-2.10	-3.06	A
maker-ptg000650l-snap-gene-20.28-mRNA-1	<i>ATP1a3</i>	sodium/potassium-ATPase $\alpha$ -subunit	sodium/potassium-transporting ATPase subunit alpha-3	-	-4.83	-4.21	A
maker-ptg001619l-snap-gene-1.72-mRNA-1	<i>NKA1/NI1</i>	regulation of sodium ion transport	sodium/potassium transporting ATPase interacting 1	-	-2.96	-1.98	A
maker-ptg000175l-snap-gene-20.31-mRNA-1	<i>CACNA1E</i>	voltage-gated calcium channel	calcium voltage-gated channel subunit alpha1E	-	4.84	4.33	A
maker-ptg000810l-snap-gene-9.26-mRNA-1	<i>TRPC6</i>	receptor-activated calcium channel	short transient receptor potential channel 6	-	-1.54	-1.00	A
maker-ptg000049l-augustus-gene-2.28-mRNA-1	<i>P2RX5</i>	ligand-gated ion channel	P2X purinoceptor 5	-	5.13	2.46	A
maker-ptg000384l-snap-gene-5.22-mRNA-1	<i>GLRA1</i>	ligand-gated chloride channel	glycine receptor alpha 1	-	-4.78	-5.45	A
snap_masked-ptg000276l-processed-gene-2.10-mRNA-1	<i>MYL4</i>	regulatory light chain of myosin	myosin light chain 4	4.80	-	-	B
snap_masked-ptg002519l-processed-gene-0.5-mRNA-1	<i>ACTC1b</i>	actin isoform, striated muscle	actin, alpha cardiac muscle 1b	2.54	-	-	B
maker-ptg000333l-augustus-gene-6.128-mRNA-1	<i>MYBPC3</i>	regulator of cardiac contraction	myosin-binding protein C, cardiac-type	-3.19	-	-	B
maker-ptg001199l-augustus-gene-0.2-mRNA-1	<i>PP1L6</i>	protein folding and protein peptidyl-prolyl isomerization in cytoplasm	peptidylprolyl isomerase like 6	5.95	-	-	B
maker-ptg000775l-snap-gene-5.52-mRNA-1	<i>TRIM35</i>	protein ubiquitination	E3 ubiquitin-protein ligase	5.35	-	-	B
maker-ptg000650l-snap-gene-19.18-mRNA-1	<i>CEACA1</i>	cell-cell adhesion molecule	carcinoembryonic antigen-related cell adhesion molecule 1	5.08	-	-	B
snap_masked-ptg001288l-processed-gene-5.53-mRNA-1	<i>PP1C</i>	assists protein folding	peptidylprolyl isomerase C	-5.01	-	-	B
maker-ptg000041l-snap-gene-15.5-mRNA-1	<i>FKBP1a</i>	protein folding and trafficking	peptidyl-prolyl cis-trans isomerase	-5.76	-	-	B
maker-ptg001424l-augustus-gene-5.112-mRNA-1	<i>TSPAN1</i>	protein maturation	tetraspanin 15	5.20	-	-	B

maker-ptg0002431-augustus-gene-0.26-mRNA-1	<i>NXPE2</i>	integral component of membrane	neurexophilin and PC-esterase domain family member 2	6.37	-	-	other	B
maker-ptg0008521-snap-gene-27.43-mRNA-1	<i>CCL4</i>	cell-cell signaling	C-C motif chemokine 4 homolog	-6.88	-	-	signaling	B
maker-ptg0000511-snap-gene-32.5-mRNA-1	<i>ZBED4</i>	transcriptional regulator	zinc finger BED domain-containing protein 4	5.19	-	-	transcription factor	B
maker-ptg0002381-snap-gene-24.29-mRNA-1	<i>GTF2IR D2</i>	DNA-binding transcription factor activity	general transcription factor II-I repeat domain-containing protein 2	6.94	-	-	transcription factor	B
snap_masked-ptg0024621-processed-gene-0.59-mRNA-1	<i>HES5</i>	transcription factor; regulates cell differentiation	transcription factor HES-5	5.54	-	-	transcription factor	B
maker-ptg0004241-snap-gene-12.1-mRNA-1	<i>TCF15</i>	transcription factor; regulates patterning of axial skeleton and skeletal muscles	transcription factor 15	-4.31	-	-	transcription factor	B
maker-ptg0003121-snap-gene-4.182-mRNA-1	<i>ABCC1</i>	ATP-binding cassette transporter	multidrug resistance protein 1	5.78	-	-	transmembrane ion transport	B
maker-ptg0001601-snap-gene-11.44-mRNA-1	<i>CHRND</i>	opening of an ion-conducting channel across the plasma membrane	acetylcholine receptor subunit delta	1.08	-	-	transmembrane ion transport	B
maker-ptg0005771-augustus-gene-2.33-mRNA-1	<i>CACNA1d</i>	mediates the entry of calcium ions into excitable cells	voltage-dependent L-type calcium channel subunit alpha-1D	4.29	-	-	transmembrane ion transport	B
maker-ptg0018441-augustus-gene-4.132-mRNA-1	<i>KCNV1</i>	down-regulate the channel activity of <i>KCNB1</i> , <i>KCNB2</i> , <i>KCNK4</i> and <i>KCND1</i> , possibly by trapping them in intracellular membranes	potassium voltage-gated channel modifier subfamily V member 1	4.58	-	-	transmembrane ion transport	B
snap_masked-ptg0003161-processed-gene-6.4-mRNA-1	<i>KCNH1</i>	delayed rectifier potassium channel	potassium voltage-gated channel subfamily H member 1	1.77	-	-	transmembrane ion transport	B
maker-ptg0027901-snap-gene-0.5-mRNA-1	<i>KCNQ4</i>	regulation of neuronal excitability	potassium voltage-gated channel subfamily Q member 4	1.30	-	-	transmembrane ion transport	B
snap_masked-ptg0010911-processed-gene-3.28-mRNA-1	<i>KCNB2</i>	delayed-rectifier voltage-dependent outward potassium current	potassium voltage-gated channel subfamily B member 2	-2.57	-	-	transmembrane ion transport	B
snap_masked-ptg0006331-processed-gene-26.11-mRNA-1	<i>KCNA6a</i>	voltage-gated potassium channel	shaker-related potassium channel tsha2	-3.39	-	-	transmembrane ion transport	B
maker-ptg0015871-snap-gene-9.16-mRNA-1	<i>SLC6a17</i>	sodium-dependent neutral amino acid transporter	solute carrier family 6 member 17	5.45	-	-	transmembrane ion transport	B
maker-ptg0001021-augustus-gene-69.90-mRNA-1	<i>SLC20a1b</i>	sodium-dependent phosphate transporter 1-B	solute carrier family 20 member 1b	1.04	-	-	transmembrane ion transport	B
maker-ptg0025191-snap-gene-5.142-mRNA-1	<i>SLC10a1</i>	sodium/bile acid cotransporter	solute carrier family 10 member 1	-2.87	-	-	transmembrane ion transport	B
maker-ptg0015741-snap-gene-0.9-mRNA-1	<i>SLC8a1</i>	sodium/calcium exchanger 1	solute carrier family 8 member 1	2.81	-	-	transmembrane ion transport	B
maker-ptg0014511-snap-gene-2.178-mRNA-1	<i>CACNA1c</i>	mediate the influx of calcium ions into the cell upon membrane polarization	calcium voltage-gated channel subunit alpha1 C	-3.61	-	-	transmembrane ion transport	B
maker-ptg0004571-augustus-gene-12.7-mRNA-1	<i>ACTN2</i>	F-actin cross-linking protein	alpha-actinin-2	4.76	3.71	6.17	cytoskeletal & sarcomeric	C

maker-ptg000731l-augustus-gene-2.84-mRNA-1	<i>MYO1f</i>	unconventional myosin; vesicle transport along actin filaments	unconventional myosin-If	1.25	1.58	2.44	C	cytoskeletal & sarcomeric
maker-ptg000407l-augustus-gene-3.0-mRNA-1	<i>MYH7</i>	myosin heavy chain beta isoform	myosin-7	-5.05	-4.23	-4.09	C	cytoskeletal & sarcomeric
maker-ptg000253l-augustus-gene-66.24-mRNA-1	<i>Ca14</i>	type I membrane protein; regulation of pH	carbonic anhydrase 14	-1.95	-6.40	-3.81	C	other
maker-ptg001301l-snap-gene-0.33-mRNA-1	<i>SLC7a3</i>	sodium ions- and pH-independent amino acid transport	solute carrier family 7 member 3	-1.64	-1.61	-1.34	C	other
maker-ptg000102l-snap-gene-29.0-mRNA-1	<i>SUB1</i>	may be involved in stabilizing the multiprotein transcription complex	activated RNA polymerase II transcriptional coactivator p15	2.30	6.07	4.13	C	other
maker-ptg000939l-snap-gene-2.46-mRNA-1	<i>POSTN</i>	extracellular matrix organization	periostin	-7.09	-2.71	-3.06	C	other
maker-ptg000159l-snap-gene-9.79-mRNA-1	<i>MMP13</i>	degradation of extracellular matrix proteins	collagenase 3	-6.47	-4.67	-2.33	C	other
maker-ptg000065l-augustus-gene-13.18-mRNA-1	<i>PCOLCE2</i>	collagen biosynthesis and modifying enzyme	procollagen C-endopeptidase enhancer 2	-4.10	-5.29	-4.56	C	other
maker-ptg000067l-augustus-gene-17.208-mRNA-1	<i>PAPLN</i>	extracellular matrix organization	papilin	-4.80	-2.87	-6.51	C	other
maker-ptg001300l-augustus-gene-3.31-mRNA-1	<i>ENTPD5</i>	mediate catabolism of extracellular nucleotides	ectonucleoside triphosphate diphosphohydrolase 5	-4.32	-5.44	-6.59	C	other
maker-ptg000056l-snap-gene-10.13-mRNA-1	<i>ADH1</i>	ethanol catabolic process	alcohol dehydrogenase 1	-2.75	-6.78	-7.33	C	other
maker-ptg000435l-snap-gene-1.208-mRNA-1	<i>GPR22</i>	G-protein coupled receptor	probable G-protein coupled receptor 22	4.31	3.92	5.86	C	signaling
maker-ptg000574l-snap-gene-0.16-mRNA-1	<i>CABP2</i>	signal transduction	calcium-binding protein 2	-1.41	-4.42	-6.12	C	signaling
maker-ptg000393l-snap-gene-34.55-mRNA-1	<i>PHF24</i>	regulate G protein-coupled receptor signaling pathway	PHD finger protein 24	-3.16	-3.99	-4.96	C	signaling
snap_masked-ptg000659l-processed-gene-2.105-mRNA-1	<i>SLC4a5</i>	electrogenic sodium bicarbonate cotransporter 4	solute carrier family 4 member 5	6.84	3.48	3.44	C	transmembrane ion transport
maker-ptg000509l-snap-gene-9.39-mRNA-1	<i>ATP1b1b</i>	potassium/sodium-ATPase $\beta$ -subunit	ATPase potassium/sodium transporting beta 1b	1.48	3.05	1.98	C	transmembrane ion transport
maker-ptg000073l-augustus-gene-19.8-mRNA-1	<i>KCNH2</i>	voltage-gated inwardly rectifying potassium channel	potassium voltage-gated channel subfamily H member 2	-1.29	-2.79	-2.60	C	transmembrane ion transport
maker-ptg001156l-snap-gene-2.33-mRNA-1	<i>KCNJ10</i>	inward rectifier-type potassium channel	potassium voltage-gated channel subfamily J member 10	-2.04	-5.33	-3.78	C	transmembrane ion transport

## Supplementary Table 2

**Supplementary Table 2** Significantly enriched Gene Ontology (GO) terms in set A genes with Fisher's exact test p-value < 0.05.

Term	GO terms	Category	Count	%	P-value	Genes	List Total	Pop Total	Pop Hits	Fold Enrichment	Benja ironi mini	FDR	
GO:0048246	macrophage chemotaxis	Biological Process	7	0.89	4.68E-06	P2RX7, WASB, CXCR3.2, IL1B, RAC2, DUOX, CSF1RA	656	18397	14	14.022	0.006	0.006	
GO:0005886	plasma membrane	Cellular Component	156	19.7	4.83E-05	NRROS, IGF2BP2A, CPNE7, CSEL1, ITPR1B, INPP5KB, CELSR2, CLDN1, ZGC:162193, RAB44, GJA3, S1PR5A, TSPAN4A, DUOX, PLYXNC1, MFS2D2B, SLC39A1, FZD9B, CA14, PHLDA3, TESCA, ENTPD2A.2, KCNH3, TESCB, CLDND, ENTPD2A.1, ENTPD3, SEMA6E, NRP2A, ADGRA3, GABRG2, APBA1A, MAG, GPR160, AMPH, ALPK2, NLGN4XA, RHOBTB2B, ABCG1, SLC6A19A.1, GRIA3B, SKAP2, ABHD2A, ESYT1B, PCDH11, SHC1, RGS7BPB, S1:CH211-20MC21.1, KIRREL3L, RASGRP4, RASGRP3, NKAIN1, RAPGEF5A, CXCR3.2, SCUBE1, PIP5K1CA, CCDC141, CNRIP1B, TSPAN18A, FZD1, HSPA8, SYT3, CARMIL2, GAS7A, TLN2B, S1:CH73-206P6.1, APCDD1L, ADGRB1A, SNX18A, JAG2B, FYBB, ADI1, BAMBIA, RGL1, MPP2B, TRPC6A, PLEK, PIK3CD, NRCAMA, TRH, CSF1RA, RND2, MRC1A, ADRB3A, MMP24, CHRNA2B, CTSK, RAC2, EFN2A, S1:CH211-132E22.4, DCHS1B, HEPHL1B, AMIN, TRPC3, KREMEN1, CALCLA, LSR, IRAK3, OPRM1, TNFRSF1B, GMIP, AGTR1A, ADGRA1A, RGMA, KCN2A, BIN2B, SLC15A1A, SLC4A2B, ZGC:65811, LCP1, CDH17, OBSCNB, GPSM1B, NEO1A, PCDH10A, GUCY2F, NKD1, ADCY3A, RHD, SEMA3AB, S1:CH1073-396H14.1, ATP1A3A, GLRA1, CNR1, RYR2B, RAPIAB, SLC2A1B, ZGC:101731, CNNM3, RASGEF1BA, TSPAN2A, CAP2, PDZD7A, MYO1EA, IL10RA, MERTKA, STAC3, PCDHB, MST1RB, S1:DKY-11F4.7, ADGRL2B.1, SNX9B, P2RX7, GHRA, TSPAN14, SEMA4BA, P2RX5, ANO8B, AXL, AGRN, EFR3BB, MXG, EGFRA, LGR4, CRFB4, GRNA	615	739	17340	1.831	0.052	0.054	0.054
GO:0005509	calcium ion binding	Molecular Function	48	6.08	7.66E-05	CNTNAP1, ESYT1B, MYL18B, PCDH11, ITPR1B, LTBP1, NELL2B, CELSR2, PCDH10A, NKD1, RASGRP4, RASGRP3, PRF1.8, MEGF6A, SCUBE1, EGFLAM, RYR2B, PLCD3A, DUOX, FBN2B, DCHS1B, TESCA, TESCB, SYT3, PCDH9, EPDR1, TBC1D9, SLC25A23B, PCDHB, AIF1L, VIL1, JAG2B, SMOCC2, LOXL2A, LOXL2B, NCALDB, PLCH1, MASP2, KCNIP1B, EFHC2, LCP1, AGRN, PVALB3, CDH17, PLA2G4AA, MACF1A, PLA2G4AB, CDH19	656	5	18397	22.435	0.442	0.198	0.198
GO:0045921	positive regulation of exocytosis	Biological Process	4	0.51	4.26E-04	CADPSA, CADPS2, CADPSB, RAB27A	656	78	18397	3.955	0.448	0.198	
GO:0019221	cytokine-mediated signaling pathway	Biological Process	11	1.39	4.34E-04	GHRA, IGF2BP2A, IL2R2B, IL10RA, IL1B, EB13, STAT4, IRAK3, CSF1RA, IL6R, CRFB4	656	73	18397	3.842	0.781	0.379	
GO:0030593	neutrophil chemotaxis	Biological Process	10	1.27	0.001105	LYN, WASB, SAA, CCL19B, IL1B, RAC2, DUOX, LTA4H, CCL3A.4, CCL35.2	656	7	18397	16.025	0.856	0.388	
GO:0014005	microglia development	Biological Process	4	0.51	0.001412	NRROS, SLC7A7, IRF8, CSF1RA	615	220	17340	2.307	0.786	0.770	
GO:0005216	ion channel activity	Molecular Function	18	2.28	0.002204	KCNH3, TRPC6A, TRPC3, KCNC2, ITPR1B, KCNC1A, GABRG2, P2RX7, GLRA1, P2RX5, TRPN1, RYR2B, CHRNA2B, KCNO5A, CACNA1DA, CACNA1EB, KCNS3A, GRIA3B	656	130	18397	2.804	0.961	0.539	
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	Biological Process	13	1.65	0.002356	LYN, SHC1, TNK1, MERTKA, CSF1RA, SLA1A, PTK2BB, MST1RB, DOK2, AXL, BLNK, PTK6A, EGFRA	663	71	18868	3.607	0.581	0.435	
GO:0030027	lamellipodium	Cellular Component	9	1.14	0.003344	VIL1, CARMIL2, CAPGB, NAV3, PHACTR4A, ABI3A, IGF2BP1, ILK, ARHGEF7B	656	3	18397	28.044	0.994	0.727	
GO:1990504	dense core granule exocytosis	Biological Process	3	0.38	0.003707	CADPSA, CADPS2, CADPSB	656	3	18397	28.044	0.994	0.727	



GO:0005085	guanyl-nucleotide exchange factor activity	17	2.15	0.004239	FARP2, OBSCNB, ARHGGEF33, ARHGGEF12B, ARHGGEF7B, RASGRP4, RASGRP3, RAPGEF5A, SI:DKEY-65J6.2, ARHGGEF10LB, CCDC88C, CYTH3B, ARHGGEF25A, SI:DKEY-33C9.6, RGL1, MCF2L2, RASGEF18A	615	215	17340	2.229	0.949	0.815	0.815
GO:0007264	small GTPase mediated signal transduction	11	1.39	0.004526	ARHGAP32A, RAPGEF5A, GDI1, CCDC88C, RAC2, RGL1, RND2, RASGEF18A, RHOB1B2B, RASGRP4, RASGRP3	656	106	18397	2.910	0.998	0.745	0.745
GO:0004896	cytokine receptor activity	8	1.01	0.004663	GHRA, IGF2BP2A, IL2RGG, IL10RA, EBI3, CD74A, IL6R, CTRF4	615	59	17340	3.823	0.962	0.815	0.815
GO:0030154	cell differentiation	30	3.8	0.004889	VEGFAA, SPI1B, SLA1A, PTK2BB, FGF4, FSTL3, IFNG, SEMA3AB, EFN2A, FSTA, PTK6A, LYN, SRRM4, DLX1A, TNK1, NRP2A, NRG1, TDRD5, RNASEL3, FGF1A, IFT20, SMAD6B, ERF, ID4, PTPN6, TRIP13, AGRN, LGR4, EGFR, RARGA	656	487	18397	1.728	0.999	0.745	0.745
GO:0031234	extrinsic component of cytoplasmic side of plasma membrane	7	0.89	0.005174	LYN, ESYT1B, TNK1, STAC3, PTK6A, SLA1A, PTK2BB	663	46	18868	4.331	0.740	0.448	0.448
GO:0008201	heparin binding	9	1.14	0.007297	VEGFAA, SMOC2, VEGFBA, SI:DKEY-6N6.1, SI:CH211-106H11.3, RSPO3, NRP2A, NELL2B, FGF1A	615	80	17340	3.172	0.994	1.000	1.000
GO:0005737	cytoplasm	182	23	0.008406	CYFIP2, TRIM35-12, ANKRD37, NCF1, IGF2BP2A, IRF1B, CSE1L, SI:CH211-163L21.7, CPOX, TRAF4A, INPP5KB, NADSYN1, NELL2B, RCBTB1, FGF4, SELENOU1A, S1PR5A, TRIM35-19, PIM3, NTF5C1AA, EGLN1A, ZMP:0000000936, EGLN1B, PHLDA3, ANKS1B, TESCA, TPD52, TESCB, FTHL27, CDK15, NEFMA, YOD1, VASH1, NCCRP1, MID1, APBA1A, MARF1, M17, DOK2, SI:CH211-253B8.5, SI:CH211-256M1.8, MAK, HHIPL1, AMPH, PFN1, OPTN, ALOX5A, SKAP2, ILRUN, SI:DKEY-33C12.3, PHACTR4A, TRIM35-36, UBA7, SHC1, ARHGFE12B, RGS7BPB, TUBA8L2, CITA, SI:CH211-204C21.1, IQGAP3, SUL1T1S3, PRDX5, CSRP1B, NUAK2, KIF3B, PCBP4, INPP5D, CASKIN1, SURF6, PLCD3A, LTA4H, SRGAP3, LHPP, EGR1, HSPA8, CARMIL2, ZGC:92594, GAS7A, TLM2B, TNK1, PLK1, SI:CH211-108D22.2, TDRD5, SI:CH211-266G18.9, FGF1A, HPR1L, GFAP, IFT20, VIL1, IRAK1BP1, SMAD6B, ADI1, PPP1R1C, CCDC88C, MYH7L, MYO5C, ID4, DUSP22A, RPS21, SI:DKEY-23E4.3, MACF1A, GALK1, BTG1, MYL1PB, GDI1, LRRC39, ILK, PIK3CD, SLA1A, DUSP16, TUBA1C, UCHL1, AIFM2, NMRK1, RAC2, IP6K2A, MCF2L2, UNKL, DUSP2, PLEKHO1B, GSTO1, PABPC4, COP22, IRAK3, SI:DKEY-37M8.11, APRT, GCHFR, VWA8, BIN2B, TRAF5, LCP1, ACSBG2, PLA2G4AA, SI:DKEY-1K23.3, RBM22, PLA2G4AB, RPRMA, CAPGB, DNAH2, TPMT.1, DNAH5, NUAK1A, PPM1H, NOD2, CD74A, CRIP1, NKD1, DNAJB2, ABLIM3, CNR1, IGF2BP1, STAT4, CAP2, ATE1, PRKAG2A, PTPN18, URGCP, MYOIEA, FGF13A, STAC3, FKBP1AB, MYH11A, ELP3, YWHABL, PPP3CCA, TUBB4B, ARID3C, ARHGFE7B, PLEKHA8, CREBBPB, SI:CH211-195H23.3, GADD45AB, CAPN10, PTPN6, TBKBP1, FBXO40.1, PVALB3, RGS14B, MXG, CAMK1DA, NEK12	663	4405	18868	1.176	0.889	0.491	0.491
GO:0003779	actin binding	22	2.78	0.009035	MICAL2B, CAPGB, PDLIM5B, PHACTR4A, TLM2B, MYOIEA, MYH11A, MTSS1LB, VIL1, MTSS1LA, WASB, PDLIM2, NAV3, ABLIM3, MYO23A, NCALDB, MYH7L, MYO5C, LCP1, PFN1, MACF1A, CAP2	615	337	17340	1.841	0.998	1.000	1.000
GO:0009925	basal plasma membrane	4	0.51	0.009451	CSE1L, CACNA1DA, MST1RB, EGFR	663	13	18868	8.756	0.915	0.491	0.491
GO:0001935	endothelial cell proliferation	4	0.51	0.009837	LOXL2A, LOXL2B, ITGB8, ARHGFE7B	656	13	18397	8.629	1.000	1.000	1.000
GO:0097178	ruffle assembly	3	0.38	0.011783	MTSS1LA, MTSS1LB, AIF1L	656	5	18397	16.827	1.000	1.000	1.000
GO:0050832	defense response to fungus	3	0.38	0.011783	NCF1, RAC2, DUOX	656	5	18397	16.827	1.000	1.000	1.000

GO:0001965	G-protein alpha-subunit binding	Molecular Function	4	0.51	0.012012	GPSM1B, CDC88C, OPRM1, RGS14B	615	14	17340	8.056	1.000	1.000	1.000
GO:0098609	cell-cell adhesion	Biological Process	13	1.65	0.012109	OBSNCB, ZMP:0000001082, LAMA5, NEO1A, TLN2B, NRCAMA, SI:CH73-208G10.1, CELSR2, KIRREL3L, TMEEM47, SI:CH211-66E2.3, CCDC141, CDH17	656	160	18397	2.279	1.000	1.000	1.000
GO:0006629	lipid metabolic process	Biological Process	21	2.66	0.013165	SI:CH1073-429110.1, OLAH, ABHD2A, PTGIS, CERS6, ELOVL2, CSFIRA, CPT1B, SREBF2, SULT1T3, FA2H, PLD1A, ELOVL4A, FADS6, PCTY1BA, PLCD3A, PLCH1, CHPT1, PLA2G4AA, PLA2G4AB, B4GALT5	656	326	18397	1.807	1.000	1.000	1.000
GO:000902	cell morphogenesis	Biological Process	6	0.76	0.014647	DCH51B, CVFIP2, NEO1A, RGMA, CAP2, CDH19	656	41	18397	4.104	1.000	1.000	1.000
GO:0005200	structural constituent of cytoskeleton	Molecular Function	7	0.89	0.019095	TUBA1C, SI:DKEY-33C12.3, TLN2B, NEFMA, TUBA8L2, TUBB4B, GFAP	615	60	17340	3.289	1.000	1.000	1.000
GO:0000278	mitotic cell cycle	Biological Process	9	1.14	0.019973	SNX9B, TUBA1C, PLK1, TUBA8L2, PTPN6, E2F4, SNX18A, TUBB4B, POLE	656	95	18397	2.657	1.000	1.000	1.000
GO:0035335	peptidyl-tyrosine dephosphorylation	Biological Process	5	0.63	0.020926	PTPN18, DUSP2, PTPN6, TNS2A, DUSP16	656	30	18397	4.674	1.000	1.000	1.000
GO:0009952	anterior/posterior pattern specification	Biological Process	9	1.14	0.021135	HES6, HOXA4A, ZIC3, NOTUM1A, HNF1BA, EFN2A, MAFBA, AGRN, HOXD3A	656	96	18397	2.629	1.000	1.000	1.000
GO:0007399	nervous system development	Biological Process	17	2.15	0.021231	SRRM4, NEO1A, PHACTR4A, IGF2BP2A, FGF13A, MERTKA, NRG1, NRP2A, ELP3, MST1RB, LFNG, SEMA3AB, NAV3, AXL, IGF2BP1, CSPG5A, EFN2A	656	256	18397	1.862	1.000	1.000	1.000
GO:0009986	cell surface	Cellular Component	12	1.52	0.021797	ROBO2, NRROS, NEO1A, SCUBE1, LRFN1, TNFSF13, ITGB8, TMX3B, ITGB6, PDIA8, NLGN4XA, CD74A	663	156	18868	2.189	0.997	0.945	0.945
GO:0002376	immune system process	Biological Process	10	1.27	0.022778	ILRUN, CXCR3.2, IRF1B, INPP5D, IRF8, MHC2A, IFI30, CSF1RA, RNASEL3, C7B	656	116	18397	2.418	1.000	1.000	1.000
GO:0006955	immune response	Biological Process	19	2.41	0.023596	NRROS, CCR12A, CCL19B, TNFSF13, MHC2A, CD74A, CCL35.2, C7B, TNFSF13B, PRF1.8, CXCR3.2, IL6, IRAK1BP1, ZGC:174904, MHC2DEB, CTSK, IL1B, SI:CH211-214K5.3, CCL34A.4	656	302	18397	1.764	1.000	1.000	1.000
GO:0070679	inositol 1,4,5 trisphosphate binding	Molecular Function	4	0.51	0.024251	TRPC6A, TRPC3, PLCL2, ITPR1B	615	18	17340	6.266	1.000	1.000	1.000
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	Biological Process	4	0.51	0.024604	VEGFAA, GHRA, PTPRC, CD74A	656	18	18397	6.232	1.000	1.000	1.000
GO:0009617	response to bacterium	Biological Process	8	1.01	0.026753	CXCR3.2, SAA, IL1B, DUOX, LTA4H, NOD2, CPT1B, NITR3C	656	82	18397	2.736	1.000	1.000	1.000
GO:0001525	angiogenesis	Biological Process	14	1.77	0.028077	VEGFAA, SHC1, KREMEN1, CALCRLA, NRP2A, ISM1, ARHGFEF7B, RNASEL3, PTK2BB, FGF1A, PLD1A, ANGPT12B, EFN2A, ITGB8	656	201	18397	1.953	1.000	1.000	1.000
GO:0035091	phosphatidylinositol binding	Molecular Function	9	1.14	0.028296	SNX9B, ARHGAP32A, PLD1A, ESYT1B, PXDC1B, NCF1, TOM1, ITPR1B, SNX18A	615	102	17340	2.488	1.000	1.000	1.000
GO:0035556	intracellular signal transduction	Biological Process	30	3.8	0.028431	CAB39L, SHC1, PLEK, NUAK1A, NOD2, CITTA, GUCY2F, ADCY3A, RASGRP4, RASGRP3, NUAK2, BLNK, PLCD3A, MCF2L2, GPR155A, PLCL2, STAC3, PKN1A, IRAK3, NRG1, TNS2A, SI:CH211-108D22.2, GMIP, ARHGFEF7B, ASB13A.2, PPP1R1C, IMAK, PLCH1, PTPN6, RGS14B	656	560	18397	1.502	1.000	1.000	1.000
GO:0043547	positive regulation of GTPase activity	Biological Process	10	1.27	0.028944	RAPGEF5A, ADCYAP1A, CCL19B, SI:DKEY-33C9.6, RGL1, CCL34A.4, RASGEF1BA, RASGRP4, CCL35.2, RASGRP3	656	121	18397	2.318	1.000	1.000	1.000
GO:0005576	extracellular region	Cellular Component	50	6.33	0.029079	COL8A1A, CCL19B, IL19L, TRH, ISM1, IFI30, HAPLN4, FGF4, PRF1.8, CRHBP, CTSK, RSPO3, FBN2B, CCL34A.4, EPDR1, RNASEL3, LG12A, SAA, RBP4, IL1B, ZGC:65811, SMPDL3B, VEGFAA, ITIH5, LAMA5, NOTUM1A, LTBP1, CCL35.2, SEMA3AB, DIPK2AB, EGFAM, ADCYAP1A, CPAMD8, ADAMTS17, FSTA, PKDCCB, CPA4, ADAMTS15A, FGF13A, FGF1A, A2ML, C7B, SMOCC2, GDF9, IL6, LOXL2A, LOXL2B, APOC1, SI:CH211-106H11.3, GRNA	663	1059	18868	1.344	1.000	0.972	0.972
GO:0043020	NADPH oxidase complex	Cellular Component	3	0.38	0.029906	NCF1, NCF2, DUOX	663	8	18868	10.672	1.000	0.972	0.972

GO:0016491	oxidoreductase activity	34	4.3	0.030496	ACADVL, MICAL2B, CPOX, IFI30, CRYZ, SELENOUIA, HSD11B2, CYP2K21, PRDX5, AIFM2, TYRBP1B, DUOX, GLUD1B, EGLN1A, Si:DKEY-12E7.4, EGLN1B, Si:CH1073-42910.1, HEPHL1B, CYB5A, AKR1B1.2, GSTO1, SQR, YWHABL, PRDX6, FA2H, CYP26A1, LOXL2A, LOXL2B, ADI1, DNAJC10, CRYZL1, CRYL1, MAO, ALOX5A	615	661	17340	1.450	1.000	1.000	1.000	1.000
GO:0005096	GTPase activator activity	16	2.03	0.030754	RG51B, ARHGAP32A, GDI1, ARHGEF12B, TBC1D9, GMIP, IQGAP3, RASGRP3, ARHGAP22, RGS4, ARHGAP10, Si:DKEY-191M6.4, TBC1D7, Si:DKEY-33C9.6, RGS14B, Si:DKEYP-23E4.3	615	248	17340	1.819	1.000	1.000	1.000	1.000
GO:0009611	response to wounding	4	0.51	0.032621	LYN, RAC2, DUOX, PLA2G4AA	656	20	18397	5.609	1.000	1.000	1.000	1.000
GO:0007409	axonogenesis	8	1.01	0.033549	RG54, SRRM4, CNR1, Si:CH211-15918.4, ELP3, AGRN, ALDOAA, OPTN	656	86	18397	2.609	1.000	1.000	1.000	1.000
GO:0004715	non-membrane spanning protein	6	0.76	0.033744	LYN, TNK1, PTK6A, SLA1A, PTK2BB, PKDCCB	615	51	17340	3.317	1.000	1.000	1.000	1.000
GO:0019955	tyrosine kinase activity	5	0.63	0.034212	GHRA, IL2RGB, EBI3, CSF1RA, IL6R	615	35	17340	4.028	1.000	1.000	1.000	1.000
GO:0004713	protein tyrosine kinase activity	10	1.27	0.035113	LYN, TNK1, MERTKA, CSF1RA, PTK6A, CITA, SLA1A, PTK2BB, MST1RB, EGFR	615	126	17340	2.238	1.000	1.000	1.000	1.000
GO:0016020	membrane	273	34.6	0.036941	SIC35B2, IL2RGB, EXT1C, MHC2A, ITPR1B, Si:CH211-15918.4, CLDN1, ZGC:174904, GIA3, DUOX, MFSO2B, ADAM8B, ADRA1AA, ENTDP2A-2, ZMP:000001082, CLDND, TMEM1170A, ENTDP2A.1, ENTDP3, Si:DKEY-26C10.5, ADGRA3, NRP2A, ATP6AP1A, CPT1B, KCNCA1A, TMEM1170A, GABRG2, GPR160, ALPK2, SLC6A19A.1, ESYT1B, PCDH11, PEX11A, ARHGEF12B, EBI3, RGS7BPB, Si:CH73-171020.1, XKRX, HSD11B2, NKAIN1, INPP5D, CACNA1EB, CSPG5A, ST3GAL4, PLCD3A, LINGO1A, TSPAN18A, PRR11, ABCC6A, FZD1, CARMIL2, TMEM86B, ST8SIA3, TNFSF13, APCDD1L, RAB27A, ADGRB1A, B4GALNT4A, PONZR1, SNX18A, JAG2B, CYP26A1, FA2H, TRPN1, ELOVL4A, LOXL2A, LOXL2B, ADI1, CACNA1DA, BAMBIA, MACF1A, OGFER, ROBO2, TRPC6A, KCNC2, PTPRO, SLC35F4, AP2A1, TMEM47, LFNG, IMMP2L, UCHL1, MMP24, CHRNA2B, CTSK, TYRPIB, EFN2A, Si:CH211-132E22.4, Si:DKEY-33C9.6, HAVCR1, DCHS1B, GPR39, PLEKHO1B, FCER1G, ELOVL2, KREMEN1, COP22, IL17RC, SREBF2, AGTR1A, RGMA, GCHFR, SLC22A23, SLC7A5, SLC7A7, IGH, ZGC:65811, VEGFAA, RPRMA, TIMM10, Si:DKEY-25110.2, CD74A, PCDH10A, GUCY2F, NKD1, ADCY3A, RHD, Si:CH1073-396H14.1, ATP1A3A, GLRA1, STAMBPL1, RYR2B, KCNQ5A, SLC2A1B, DUOX2, FRR51B, YIPF2, IL10RA, STAC3, TMEM64, NITR3C, PLEKHA8, GHRA, IL17RA1A, SEMA4BA, SC:D156, TOMM7, AGRN, EGFR, LGR4, TMEM229B, GPR146, CDC167, NRROS, IL19L, CSEIL, CPOX, ZGC:174863, CELSR2, SELENOUIA, MGAT3B, APMAP, S1PR5A, TOM1, TSPAN4A, PLXNC1, FZD9B, IL6R, SLC39A1, CA14, SYBU, ZMP:000000936, PHILDA3, GPR155A, KCNH3, CCR12A, SEMA6E, SLC25A23B, M17, MAG, TMX1, SLC27A6, HHIPL1, MAO, NUGN4X, FAM20CB, ABCG1, GRIA3B, AQPIA.1, ABHD2A, FAM49BB, NDUFB11, TMEM59L, Si:CH73-208G10.1, KIRREL3L, CXCR3.2, CYP2K21, SCUBE1, COX11, PKDCCB, SYT3, ANKRD22, DNAJC16, Si:CH211-266G18.9, IGSF11, IRAK1BP1, PTPRC, LRFN1, Si:DKEY-215K6.1, HPSE, CNTNAP1, PIK3CD, NRCAMA, CSF1RA, MRC1A, ADRB3A, TNFSF13B, MYLK4B, TMEM108, ITGB8, UGT2A2, UGT2A1, KCNS3A, ITGB6, Si:CH211-243A20.4, SLC13A1, AMN, PTGIS, CERS6, B3GNT2B, TRPC3, ABCB6A, LMF2A, CALCLA, BRINP3A.1, IRAK3, LSR, NRG1, OPRM1, SLC25A55A, TNFRSF1B, ADGRA1A, KCNJ2A, NAV3, SLC15A1A, SLC4A2B, LCPI, CDH17, B4GALT5, LAMA5, GPSM1B, NEO1A, MOSPD2, PPM1H, Si:CH211-66E2.3, CHSY1, CNR1, RAP1AB, SLC25A20, TSPAN2A, GALNT7, Si:CH1073-42910.1, CYB5A, ZGC:112965, MERTKA, PCDHB, MST1RB, Si:DKEY-11F4.7, ADGR12B.1, P2RX7, SNX9B, ZGC:194312, TSPAN14, P2RX5, PNKD, SYNGR1A, ANO8B, ZGC:123297, CHPT1, EFR3BB, MXG, CRFB4	663	7112	18868	1.092	1.000	0.981	0.981	
GO:0021754	facial nucleus development	3	0.38	0.038593	SRRM4, GJ1, CELSR2	656	9	18397	9.348	1.000	1.000	1.000	1.000

GO:0098978	glutamatergic synapse	Cellular Component	5	0.63	0.039725	CADPSA, CADPS2, CADPSB, NRP2A, ADGRA1A	663	37	18868	3.846	1.000	0.981	0.981
GO:0004867	serine-type endopeptidase inhibitor activity	Molecular Function	8	1.01	0.040373	ITIH5, SERPINB1, CPAMD8, SERPINB1L1, SERPINB14, AGRN, A2ML, PCSK1NL	615	90	17340	2.506	1.000	1.000	1.000
GO:0007155	cell adhesion	Biological Process	24	3.04	0.040467	DCHS1B, ZMP:0000001082, LAMA5, CLDND, PCDH9, PCDH11, TLN2B, Si:CH73-208G10.1, HAPLN4, CELSR2, PCDH10A, PCDHB, CLDN1, MAG, Si:CH211-106H11.3, SPP1, EFN2A, ITGB8, FBN2B, NLGN4XA, ITGB6, CCDC141, CDH17	656	437	18397	1.540	1.000	1.000	1.000
GO:0007275	multicellular organism development	Biological Process	34	4.3	0.041347	VEGFA, PHACTR4A, CSF1RA, CELSR2, FSTL3, LFN3, SEMA3AB, EFN2A, ZNF503, FSTA, FZD9B, FZD1, DCHS1B, NCOA2, HOXA4A, DLX1A, KREMEN1, MERTKA, NRP2A, ADGRA3, TDRD5, RNASEL3, FGF1A, MST1RB, JAG2B, M17, HAND2, HNF1BA, AGRN, RORAA, HOXD3A, LGR4, EGFRA, RBM22	656	675	18397	1.413	1.000	1.000	1.000
GO:0005178	integrin binding	Molecular Function	7	0.89	0.041979	ZMP:0000001082, LAMA5, TLN2B, Si:CH211-106H11.3, ITGB8, ITGB6, FGF1A	615	72	17340	2.741	1.000	1.000	1.000
GO:0007229	integrin-mediated signaling pathway	Biological Process	6	0.76	0.042565	ZMP:0000001082, LAMA5, FYBB, ILK, ITGB8, ITGB6	656	54	18397	3.116	1.000	1.000	1.000
GO:0031012	extracellular matrix	Cellular Component	14	1.77	0.044221	COL8A1A, NRROS, HAPLN4, COL19A1, MMP24, Si:DKEY-6N6.1, ADAMTS17, Si:CH211-106H11.3, Si:DKEY-65B12.6, FBN2B, HPSE, LGR4, LINGO1A, ADAMTS15A	663	218	18868	1.828	1.000	0.981	0.981
GO:0005925	focal adhesion	Cellular Component	7	0.89	0.045298	TLN2B, ILK, ITGB8, ITGB6, TNS2A, SLA1A, PTK2BB	663	74	18868	2.692	1.000	0.981	0.981
GO:0007265	Ras protein signal transduction	Biological Process	6	0.76	0.045507	RAPGEF5A, DOK2, RGL1, RASGEF1BA, RASGRP4, RASGRP3	656	55	18397	3.059	1.000	1.000	1.000
GO:0048703	embryonic viscerocranium morphogenesis	Biological Process	9	1.14	0.04667	EDARADD, EGR1, SMO2, DLX1A, ZIC2A, HAND2, FSTA, PONZR1, VGLL2A	656	112	18397	2.254	1.000	1.000	1.000
GO:0009134	nucleoside diphosphate catabolic process	Biological Process	3	0.38	0.047124	ENTPD2A.2, ENTPD2A.1, ENTPD3	656	10	18397	8.413	1.000	1.000	1.000
GO:0051252	regulation of RNA metabolic process	Biological Process	3	0.38	0.047124	IGF2BP2A, PCBP4, IGF2BP1	656	10	18397	8.413	1.000	1.000	1.000
GO:0043197	dendritic spine	Cellular Component	6	0.76	0.049155	KIF3B, Si:DKEY-33C9.6, SLA1A, PTK2BB, GRIA3B, APBA1A	663	57	18868	2.996	1.000	0.983	0.983

## Supplementary Table 3

**Supplementary Table 3** Significantly enriched Gene Ontology (GO) terms in set B genes with Fisher's exact test p-value < 0.05.

Term	GO terms	Category	Count	% P-value	Genes	List Pop		Total		Fold		Bonferroni		Benjamini		FDR
						Hits	Pop	Hits	Pop	Enrichment	ni	ni	ni			
GO:0031012	extracellular matrix	Cellular	28	2.95	4.29E-07	COL16A1, ECM2, COL11A2, ADAMTS12, HAPLN3, TSKU, COL1A1A, ADAMTS17, TIMP2B, TIMP2A, COL15A1B, CCN5, ADAMTSL7, MMP2, MMP13A, MMP15B, MMP17A, CCN2B, COL18A1A, COL2A1B, MGP, COL4A6, MMP19, COL4A5, SI:DKEY-65B12.6, LRRN1, TGFB1, VCANA	787	218	18868	3.07930476	1.29E-04	1.29E-04	1.29E-04	1.29E-04	1.29E-04	1.29E-04
GO:0030198	extracellular matrix organization	Biological Process	18	1.9	5.01E-05	COL16A1, ECM2, MMP2, COL11A2, MMP13A, MMP15B, MMP17A, ADAMTS12, COL18A1A, COL1A1A, ADAMTS17, COL2A1B, COL15A1B, COL4A6, COL4A5, MMP19, ADAMTSL7, TGFB1	784	133	18397	3.17579024	0.06752	0.06971	0.06971	0.06971	0.06971	0.06971
GO:0007411	axon guidance	Biological Process	21	2.22	9.99E-05	ROBO2, ROBO4, SEMA6A, SLIT1A, SI:DKEY-49N23.1, NRCAMA, SEMA3E, KIF5AA, ROBO1, SEMA4A, EFNB1, CNTN1A, LRTM2A, NPTNA, SEMA3FB, TRIOB, RTN4RL1B, NFASCA, KALRNA, EPHA3, EFNA3A	784	182	18397	2.70756476	0.130142	0.06971	0.06971	0.06971	0.06971	0.06971
GO:0005911	cell-cell junction	Cellular	12	1.27	1.07E-04	IGSF11, RAP1B, USP53B, EPB41L3A, PARD3AB, PERP, LIN7A, SI:CH211-186J3.6, MAG1B, SI:DKEY-11F4.20, NECTIN1B, LIMS1	787	68	18868	4.23080948	0.031581	0.01604	0.01604	0.01604	0.01604	0.01604
GO:0007155	cell adhesion	Biological Process	35	3.69	5.44E-04	PCDH18B, ROBO4, PCDH2G16, AMIGO1, NEDD9, SI:CH73-208G10.1, THY1, PTPRFB, PCDH19, HAPLN3, CELSR2, PARD3AB, CNTN1A, ITGA2.2, CCN5, CLDN11A, NFASCA, PCDH2AC, ITGB6, PCDH15B, ZMP:0000001082, EGF16, PCDH2G28, ITGA1, PKP3A, NRXN1A, CCN2B, FAP, NPTNA, CDH11, CTNNB1, TGFB1, VCANA, SI:CH211-186J3.6, NECTIN1B	787	437	18397	1.87939278	0.531959	0.253	0.253	0.253	0.253	0.253
GO:0005615	extracellular space	Cellular	68	7.17	0.00121	IL15L, COL16A1, EFEMP2B, ZGC:110239, SLIT1A, CCL38.1, SPINT2, OLFML2A, CLU, TSKU, SEMA4A, WNT2BB, CTSS2.2, COL1A1A, LIPG, ANGPTL2B, SVEP1, CCL19A.1, SEMA6A, MYOC, WNT5A, SI:DKEY-49N23.1, IGFBP5B, COL18A1A, ANGPTL7, ANGPTL6, COL6A2, COL6A1, SEMA3FB, COL4A6, MMP19, COL4A5, VWAA2, SMPDL3A, GSNA, CFD, COLEC10, COL11A2, LPL, SEMA3E, PRELP, CCL35.1, CST3, FRZB, CLEC3BA, TIMP2B, TIMP2A, COL15A1B, COL7A1L, WNT2, SFRP1A, TGFB2, ANGPT1, LGALS3B, NOG2, STC2A, ADMA, GDF7, MANF, BMP4, FGF19, COL2A1B, CPE, SI:DKEY-65B12.6, LRRN1, TGFB1, CBLN4, WNT7AA	787	1102	18868	1.47937561	0.304543	0.12099	0.12099	0.12099	0.12099	0.12099
GO:0072686	mitotic spindle	Cellular	7	0.74	0.00184	TPX2, POLN, KIFC1, EFHC1, WDR62, EML3, RTRAF	787	32	18868	5.24444092	0.423821	0.13771	0.13771	0.13771	0.13771	0.13771
GO:0060047	heart contraction	Biological Process	11	1.16	0.00205	SLC8A4A, RBFOX1, CMLC1, BAG3, TTN.2, CRIP2, CCM2, TPM4A, FBXO32, CRIP1, LIMS1	784	80	18397	3.22651467	0.942901	0.58064	0.58064	0.58064	0.58064	0.58064
GO:0050919	negative chemotaxis	Biological Process	8	0.84	0.00208	SEMA4A, SEMA6A, LRTM2A, SLIT1A, SEMA3FB, SI:DKEY-49N23.1, SEMA3E, RTN4RL1B	784	43	18397	4.36568581	0.945319	0.58064	0.58064	0.58064	0.58064	0.58064
GO:0005581	collagen trimer	Cellular	9	0.95	0.00296	COL18A1A, COLEC10, COL1A1A, COL2A1B, COL11A2, COL15A1B, COL4A6, COL4A5, COL7A1L	787	59	18868	3.6571404	0.58905	0.17759	0.17759	0.17759	0.17759	0.17759
GO:0098609	cell-cell adhesion	Biological Process	16	1.69	0.00348	ZMP:0000001082, ITGA1, TJP1A, PKP3A, NRCAMA, SI:CH73-208G10.1, CELSR2, CNTN1A, DLG3, ITGA2.2, PERP, CDH11, CTNNB1, NFASCA, SI:CH211-186J3.6, LIMS1	784	160	18397	2.34655612	0.992323	0.81017	0.81017	0.81017	0.81017	0.81017

GO:0007519	skeletal muscle tissue development	Biological Process	9	0.95	0.00615	SLC8A4A, RBM24A, BAG3, TTN.2, TTN.1, MYOD1, COL6A1, PPARGC1A, PXNA	784	65	18397	3.24907771	0.999816	1	
GO:0008201	heparin binding	Molecular Function	10	1.05	0.00653	CCN2B, ECM2, LRTM2A, SLIT1A, LIPG, CCN5, LPL, RTN4RL1B, PTN, PTPRFB	734	80	17340	2.95299728	0.986275	1	
GO:0045165	cell fate commitment	Biological Process	7	0.74	0.00827	BMP4, WNT2BB, WNT5A, GATA3, WNT7AA, WNT2, SOX5	784	42	18397	3.91092687	0.999991	1	
GO:0031290	retinal ganglion cell axon guidance	Biological Process	7	0.74	0.00827	ROBO2, FGF19, MMP2, PTCH2, CDH11, COL4A5, ADCY8	784	42	18397	3.91092687	0.999991	1	
GO:0004725	protein tyrosine phosphatase activity	Molecular Function	13	1.37	0.00853	SSH1A, PTPN1, PTPN9B, EYA1, DUSP8A, PTPRFB, PTPREA, DUSP26, CDC25B, PTPRD, PTPRIB.1, PTPN6, DUSP22B	734	129	17340	2.38071098	0.996341	1	
GO:0051015	actin filament binding	Molecular Function	20	2.11	0.00891	MYO5AA, TPM1, MYOM1A, PPP1R9A, MYO19, VIL1, TRIOBPB, MYO1B, SAMD14, LASP1, TPMA, GAS2, MARCKSA, FLNB, TPM4A, UTRN, MYH10, SI:DKFY-40C11.2, DBN1, GSNA	734	247	17340	1.91287273	0.997159	1	
GO:0045879	negative regulation of smoothed signaling pathway	Biological Process	5	0.53	0.00915	SOX11A, PTCH1, PTCH2, GATAD2B, SOX4A	784	20	18397	5.86639031	0.999997	1	
GO:0005085	guanyl-nucleotide exchange factor activity	Molecular Function	18	1.9	0.00999	VAV3B, EFF1DB, MCF2A, ARHGEF16, RAB3IP, DOCK11, PSD2, TRIOB, ARHGEF4, RAPGEF1B, ARHGEF2, ARHGEF25A, KALRNA, ECT2, MCF2L2, DOCK1, ARFGFE3, DOCK9B	734	215	17340	1.97782143	0.998603	1	
GO:0009008	DNA-methyltransferase activity	Molecular Function	3	0.32	0.01012	DNMT3AB, DNMT3AA, DNMT3BB.1	734	4	17340	17.7179837	0.998718	1	
GO:0071711	basement membrane organization	Biological Process	3	0.32	0.01025	FRAS1, COL4A6, COL4A5	784	4	18397	17.5991709	0.999999	1	
GO:0005178	integrin binding	Molecular Function	9	0.95	0.01082	CCN2B, ZMP:0000001082, EGF16, ITGA2.2, ITGA1, CCN5, ITGB6, TGFBI, THY1	734	72	17340	2.95299728	0.999197	1	
GO:0008375	acetylglucosaminyltransferase activity	Molecular Function	7	0.74	0.0124	GCNT4A, B3GNT2B, B3GLCTA, EXT1B, B3GNT5A, B3GALT4, RFNG	734	46	17340	3.59495321	0.999718	1	
GO:0007420	brain development	Biological Process	19	2	0.01343	MEIS1B, ROBO2, PYCR1B, SOX11A, ANGPT1, AMIGO1, TSC22D3, B3GNT5A, KIF14, NRCAMA, PCDH19, PELI1B, CNTN1A, FGF19, ZNF703, REL, MARCKSA, NFASCA, POU3F3B	784	238	18397	1.87330111	1	1	
GO:0005576	extracellular region	Cellular Component	60	6.33	0.01385	IL15L, SLIT1A, CCL38.1, NTN1A, HAPLN3, CLU, TSKU, PRF1.8, WNT2BB, COL1A1A, LIPG, CCN5, SVEP1, ADAMTSL7, CCL19A.1, NXPE3, GATD1, MYOC, MMP2, WNT5A, IGFBP5B, RNASEL3, EMILIN1A, COL4A6, COL4A5, UTS1, VCANA, SMPDL3A, COL11A2, LPL, DNAH9, PTN, CCL35.1, ADAMTSL2, LTBP1, FRZB, ADAMTSL17, DNAH9L, TIMP2A, WNT2, SFRP1A, TGFB2, RNLS, FGF13A, NOG2, CRIM1, STC2A, ADMA, GDF7, MANF, BMP4, CCN2B, CALUB, FGF19, COL2A1B, MGP, TGFB1, WNT7AA, PCOLCE2B	787	1059	18868	1.35833354	0.984757	0.53364	0.53364
GO:0005874	microtubule	Cellular Component	18	1.9	0.01388	POLN, NDE1, FGF13A, STARO9, KIF14, KIF5AA, CENPE, TPX2, TUBB5, TUBA1A, MIDJIP1L, MAP2, KIFC1, DNAH9L, MAP6D1, KIF2C, FSD1, MACF1A	787	226	18868	1.90948038	0.984909	0.53364	0.53364

GO:0005794	Golgi apparatus	Cellular Component	39	4.11	0.01423	FAIM2B, EXT1B, B3GNT5A, MOSPD1, GALNT18B, RFNG, ZDHHC20B, RAB32A, GCNT4A, CHSY1, ST3GAL8, GLCEB, VTI1A, CHSY3, ADAMI19B, VTI1B, SLC35A4, TRAPP1, B3GNT2B, RNF24, MYOC, YIPF2, COG5, ENTPD4, B3GAT2, SLC35E4, SURF4, B3GALT4, CLASP1B, ST6GALNAC5A, IFT20, B3GAT1A, CALUB, SI:CH211-269C21.2, PXYLP1, SH3GLB1A, ITM2BA, KDELR3, FGFR2	787	628	18868	1.48886767	0.986429	0.53364	0.53364
GO:0048843	negative regulation of axon extension involved in axon guidance	Biological Process	6	0.63	0.01552	SEMA44B, ROBO2, SEMA6A, SEMA3FB, Si:DKEY-49N23.1, SEMA3E	784	35	18397	4.02266764	1	1	1
GO:0050650	chondroitin sulfate proteoglycan biosynthetic process	Biological Process	4	0.42	0.01595	B3GAT1A, CHSY1, B3GAT2, PXYLP1	784	13	18397	7.22017268	1	1	1
GO:0008289	lipid binding	Molecular Function	14	1.48	0.01671	EPN3A, PAQR7B, PAQR5B, STARD5, STARD8, INSIG1, STARD9, RBP7B, ACBD5A, FABP2, CRABP1A, RBP5, ESRRGA, OSBP13B	734	158	17340	2.09326389	0.999984	1	1
GO:0003779	actin binding	Molecular Function	24	2.53	0.01671	SSH1A, MYO5AA, MICAL2B, CNN3B, TPM1, EVL1B, ANTXR2A, MTSS11B, VILI, EPB4113A, MYO1B, DAAM2, TPMA, MRIFAB, MYOZ1B, FLNB, TPM4A, UTRN, MYH10, Si:DKEY-40C11.2, SPTBN1, DBN1, GSNA, MACF1A	734	337	17340	1.68241981	0.999984	1	1
GO:0003707	steroid hormone receptor activity	Molecular Function	5	0.53	0.01715	PAQR7B, AR, ABHD2A, RXRGA, ESRRGA	734	24	17340	4.92166213	0.999988	1	1
GO:0098703	calcium ion import across plasma membrane	Biological Process	5	0.53	0.01754	SLC8A4A, CACNA1FB, CACNA1DB, CACNA1DA, CACNA1C	784	24	18397	4.88865859	1	1	1
GO:0005783	endoplasmic reticulum	Cellular Component	45	4.75	0.01853	LMBRL1, FKBP11, INSIG1, FAIM2B, EXT1B, CISD2, MOSPD1, AGPAT2, CLU, PIGX, ZDHHC20B, RNF145B, TMEM147, APL1B, ORMDL3, SEC23B, VKORC1, PTPN1, PDIA2, TRAPP1, ACSL2, MYOC, PNPLA7B, FKBP10B, SURF4, ACSL3B, ALG2, SYVN1, ELOVL6, MBOAT4, NCK2B, TMEM170A, MANF, RCN3, BACE2, LMBRD1, CALUB, DADI, CERS4A, PI4KB, KDELR3, DEGS2, SGK1, PLPP7A, HSD20B2	787	763	18868	1.41396647	0.996347	0.6178	0.6178
GO:0006470	protein dephosphorylation	Biological Process	13	1.37	0.02013	SSH1A, PTPN1, PTPN9B, PPM1H, DUSP8A, PPP4R2A, PTPRFB, PTPREA, CDC25B, PTPRD, PTPRJ1, PTPN6, DUSP22B	784	144	18397	2.11841872	1	1	1
GO:0060070	canonical Wnt signaling pathway	Biological Process	9	0.95	0.02032	SFRP1A, WNT2BB, FRZB, WNT5A, CTNNB1, TCF3B, FZD9B, WNT7AA, WNT2	784	80	18397	2.63987564	1	1	1
GO:0046983	protein dimerization activity	Molecular Function	16	1.69	0.02117	TCF15, HER9, ID2A, HER4-5, NPAS2, HEYL, FBXW11A, MYCH, MYOD1, MEF2AB, BHLHE40, TCF3B, MXD1, EBF3A, ZBED4, HIF1AL2	734	198	17340	1.90900834	0.999999	1	1
GO:0030574	collagen catabolic process	Biological Process	5	0.53	0.02309	MMP2, MMP13A, MMP15B, MMP19, MMP17A	784	26	18397	4.51260793	1	1	1
GO:0001649	osteoblast differentiation	Biological Process	4	0.42	0.02383	MYOC, PTCH1, PTCH2, NOG2	784	15	18397	6.25748299	1	1	1
GO:1901998	toxin transport	Biological Process	3	0.32	0.02421	ANTXR1C, ANTXR1B, ANTXR2A	784	6	18397	11.7327806	1	1	1
GO:0016199	axon midline choice point recognition	Biological Process	3	0.32	0.02421	ROBO2, ADCY8, ROBO1	784	6	18397	11.7327806	1	1	1
GO:0015629	actin cytoskeleton	Cellular Component	11	1.16	0.02558	VILI, TRIOBPB, MYO1B, SAMD14, MYO5AA, TPMA, MYOZ1B, PPP1R9A, MYO19, GSNA, MACF1A	787	118	18868	2.23491913	0.999579	0.76741	0.76741

GO:0030239	myofibril assembly	Biological Process	6	0.63	0.02647	CHRND, TNN12A-4, TTN-2, TTN-1, MEF2AB, TNNT3B	784	40	18397	3.51983418	1	1	1
GO:0007015	actin filament organization	Biological Process	13	1.37	0.02792	MYO5AA, TPM1, PPP1R9A, MYO19, TRIOBPB, MYO1B, SAAMD14, TPMA, MARCKSA, TPM4A, ARHGFE2, RND3A, DBN1	784	151	18397	2.02021388	1	1	1
GO:0030878	thyroid gland development	Biological Process	4	0.42	0.02842	JAG1B, GLIS3, NTN1A, BCL2L1	784	16	18397	5.86639031	1	1	1
GO:0005938	cell cortex	Cellular	9	0.95	0.03019	PARD3AB, LASP1, CTTNBP2, CTNNB1, ECT2, FNBP1L, RND3A, PXNA, MACF1A	787	88	18868	2.4519464	0.999899	0.8233	0.8233
GO:0003886	DNA (cytosine-5)-methyltransferase activity	Molecular Function	3	0.32	0.03253	DNMT3AB, DNMT3AA, DNMT3BB.1	734	7	17340	10.1245621	1	1	1
GO:0001945	lymph vessel development	Biological Process	5	0.53	0.0332	ACAA2, MMP2, SVEP1, HDAC9B, BMPR2B	784	29	18397	4.04578642	1	1	1
GO:0004721	phosphoprotein phosphatase activity	Molecular Function	12	1.27	0.03378	SSH1A, PTPN1, EYA1, SSU72, PPM1H, DUSP8A, PTPNG, PTPRFB, PPTC7A, PTPREA, DUSP22B, CDC25B	734	139	17340	2.03948013	1	1	1
GO:0004222	metalloendopeptidase activity	Molecular Function	11	1.16	0.03402	NLN, ZGC:174164, MMP2, ADAMTS17, MMP13A, MMP15B, MMP19, ADAMTS17, MMP17A, ADAMTS12, ADAM19B	734	122	17340	2.13003082	1	1	1
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	Molecular Function	6	0.63	0.03405	FKBP11, FKBP10B, FKBP7, FKBP6, FKBP5, PP1L6	734	43	17340	3.29636905	1	1	1
GO:0035335	peptidyl-tyrosine dephosphorylation	Biological Process	5	0.53	0.03706	PTPN1, EYA1, DUSP8A, PTPNG, PTPRFB	784	30	18397	3.91092687	1	1	1
GO:0008331	high voltage-gated calcium channel activity	Molecular Function	4	0.42	0.03823	CACNA1FB, CACNA1DB, CACNA1CA, CACNA1C	734	18	17340	5.24977293	1	1	1
GO:0004467	long-chain fatty acid-CoA ligase activity	Molecular Function	4	0.42	0.03823	ACSL2, ACSL3B, SLC27A1A, ZGC:101540	734	18	17340	5.24977293	1	1	1
GO:0061515	myeloid cell development	Biological Process	4	0.42	0.0389	RNF145B, TPMA, AK3, SMAD9	784	18	18397	5.21456916	1	1	1
GO:0005516	calmodulin binding	Molecular Function	11	1.16	0.0392	SLC8A4A, ASPM, KCNH1B, SPATA17, CNN3B, KCNQ5A, MAP6D1, MARCKSA, CACNA1C, UNC13BB, STRN3	734	125	17340	2.07891008	1	1	1
GO:0006486	protein glycosylation	Biological Process	12	1.27	0.04022	B3GAT1A, GCNT4A, B3GNT2B, ST3GAL8, B3GAT2, EXT1B, B3GNT5A, B3GALT4, ALG2, GALNT18B, POGUT3, ST6GALNAC5A	784	142	18397	1.98300517	1	1	1
GO:0007224	smoothed signaling pathway	Biological Process	6	0.63	0.04139	DISP2, GRK3, PTCH1, PTCH2, GLIS3, TNPO1	784	45	18397	3.1287415	1	1	1
GO:0044206	UMP salvage	Biological Process	3	0.32	0.04271	UCK2A, UPPP2, UPPP1	784	8	18397	8.79958546	1	1	1
GO:0048846	axon extension involved in axon guidance	Biological Process	3	0.32	0.04271	COL4A6, COL4A5, NTN1A	784	8	18397	8.79958546	1	1	1
GO:0005201	extracellular matrix structural constituent	Molecular Function	9	0.95	0.04299	COL18A1A, COL16A1, COL1A1A, COL2A1B, COL11A2, COL15A1B, COL4A6, COL4A5, PRELP	734	93	17340	2.28619144	1	1	1
GO:0008081	phosphoric diester hydrolase activity	Molecular Function	8	0.84	0.04391	PDE11A, TDP1, PDESAB, GPPD1, GDE1, PDE6A, PDE7A, SMPDL3A	734	77	17340	2.45443929	1	1	1



GO:0005737	cytoplasm	Cellular Component	205	21.6	0.04392	UGP2B, TRIM35-13, ARHGAP27L, Si:CH211-285F17.1, FNBPL1, TRIM2A, ZGC:110353, ZDHH208, TXNDC17, ZFVVE28, SAMD14, RPS6KA1, RNF111, TRIM35-21, RBFOX1, MBNL1, CAPN2A, STARD8, MAP4K3B, MAPK8IP2, TRIM35-20, TDDR7B, DDT4, MARCKSA, ZGC:65851, UCK2A, CRABP1A, SEC14L1, MAP2, ZNF703, RNF217, TRIOB, ANXA5B, SPTBN1, VAV3B, HSP70.3, MYO5AA, CARMIL3, LGALS3B, PKP3A, CDC7, SNF8, NCK2B, RNF34B, SNUPN, TPX2, BHMT, REL, MRV1, MMS19, DUSP22B, MACF1A, Si:CH211-191A24.4, Si:CH211-195B13.1, SYNM, ID2A, SH3BP5B, CSAD, TUBA1A, NUDCD1, COL1A1A, MCF2L2, LMCD1, RYBP, SSH1A, GATD1, ARG1, PAWR, TTBK1A, RBM24A, CNKSR3, ARHGEF2, ALDOB, SGK1, CRYAA, ALOX5B.1, RGS9B, GPS1, NEDD9, PRKAR2A, CCND2A, CRIPI1, CST3, TTN.2, FRZB, CYP2X9, RTRAF, PRKAG2A, DNMT3A, NDE1, DNMT3AA, DUSP8A, PKMA, Si:CH211-120G10.1, FABP2, CTNBN1, SH3GLB1A, LIMS1, ANKRD37, ZGC:162879, CLU, ALDH1L2, FBXW11A, TUBB5, CCND1, MID1IP1L, SESN1, S1PR5A, SESN2, EGLN1B, TXNIP, SSU72, VASH2, ACBD5A, CDC25B, DOK2, MRTFAB, TPMA, CMPK2, UTRN, SUL16B1, MAGI1B, KANK4, FKBP6, FKBP5, EPN3A, SHC1, AK3, PPP4R2A, HIMMR, NPAS2, RAP1B, PRKCA, Si:CH211-2683.4, PARD3AB, BAG3, CASKIN1, DNAH9L, GULP1A, SUL15A1, ECT2, MYH10, TRAPPC1, DPYDB, SMAD9, FBXO32, IFT20, VIL1, CNOT11, FGF19, CDC42BPAA, MAP6D1, PI4KB, HSPB8, ELAVL4, CDC42EP1A, PPP1R9A, LASP1, HYDIN, UPP2, UPP1, TNPO1, PXNA, BCAS3, USP44, Si:CH211-153B23.3, ARAP3, ASPM, CCNA1, RNF213B, TRAF6, KALRNA, TRIM54, DOCK1, GSN, PFKLB, HPGD, EVLB, PPM1H, MOSPD1, SRL, TP53INP2, FSD1, SEC23B, KSR1A, MAPK4, ARHGAP12B, PTPN1, FAXCA, RRM2, MCF2A, MAAA, POLN, GNB3A, FGF13A, ATP2B2, MYO19, ASS1, MAPK10, UFM1, MYO1B, GADD45AB, PTPN6, CCM2, PPII6	787	4405	18868	1.11572993	0.999999	1	
GO:0008345	larval locomotory behavior	Biological Process	4	0.42	0.04477	UGP2B, SETD5, PSD2, FBXO32	784	19	18397	4.94011815	1	1	1
GO:0070509	calcium ion import	Biological Process	4	0.42	0.04477	CACNA1FB, CACNA1DB, CACNA1DA, CACNA1C	784	19	18397	4.94011815	1	1	1
GO:0022625	cytosolic large ribosomal subunit	Cellular Component	6	0.63	0.04497	RPL7A, RPL18A, UBE2E1, RPL15, RPL18, RPL17	787	47	18868	3.06058558	0.999999	1	1
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	Biological Process	15	1.58	0.04552	ROBO2, PCDH15B, PCDH18B, ROBO4, CADM4, PCDH2G28, PCDH2G16, PCDH19, CELSR2, NPTNA, CDH11, PCDH2AC, Si:CH211-186I3.6, NECTIN1B, CADM1A	784	199	18397	1.7687609	1	1	1
GO:0098632	protein binding involved in cell-cell adhesion	Molecular Function	5	0.53	0.04909	ROBO4, CNTN1A, NPTNA, NRCAMA, NFASCA	734	33	17340	3.57939064	1	1	1
GO:0000278	mitotic cell cycle	Biological Process	9	0.95	0.04937	WDHD1, CENPE, ASPM, CENPF, TUBB5, TUBA1A, PTPN6, CLTCA, MCPH1	784	95	18397	2.22305317	1	1	1

## Supplementary Table 4

**Supplementary Table 4** Significantly enriched Gene ontology (GO) terms in set C genes with Fisher's exact test p-value < 0.05.

Term	GO terms	Category	Count	%	P-value	Genes	List	Pop	Pop	Fold	Enrichment	Bonferr	Benjami	FDR
							Total	Hits	Total	Enrichment	oni	oni	ni	
GO:0005576	extracellular region	Cellular	53	18.728	8.82E-18	COL5A2A, C1QTNF6A, WFDC1, LOXL4, MDKB, C1QTNF9, LOXA, HAPLN2, COMP, PRF1-5, NID2B, COL1A1A, COL1A1B, ENPP2, PAPINB, B2M, ADAMTS6, COL6A4A, MMP2, COL10A1A, BGNA, VWDE, CCN4B, SFRP2, SFRP5, COL14A1A, THBS2A, VCANA, IPPK, SOSTDC1A, SEMA3D, MSTNB, COL11A1B, FGF18A, FBLN1, PROCA1, POSTNA, FSTL5, C4, PCOLCEB, CD9A, ADAMTS17, CILP2, PRG4B, SPON2B, LGI3, COL27A1B, FIBINB, CCN2A, COL5A1, LFT1, MMP13B, GRNB	238	1059	18868	3.96760857	1.16E-15	1.16E-15	1.11E-15	
GO:0030198	extracellular matrix organization	Biological Process	21	7.4205	2.80E-16	COL6A4A, SCARA3, COL16A1, COL5A2A, COL10A1A, MMP2, COL11A1B, COL27A1B, FBLN1, MMP16B, MMP17B, MMP17A, POSTNA, COL5A1, MMP11A, COL1A1A, MMP13B, ADAMTS17, COL1A1B, PAPINB, ADAMTS6	231	133	18397	12.5748462	2.21E-13	1.86E-13	1.84E-13	
GO:0031012	extracellular matrix	Cellular	25	8.8339	4.53E-16	COL16A1, COL5A2A, COL11A1B, MMP16B, HAPLN2, POSTNA, BMPER, COL1A1A, ADAMTS17, COL1A1B, PAPINB, ADAMTS6, SCARA3, SPON2B, COL10A1A, MMP2, COL27A1B, MMP17B, MMP17A, CCN2A, CCN4B, COL5A1, MMP11A, MMP13B, VCANA	238	218	18868	9.09143474	5.86E-14	2.99E-14	2.86E-14	
GO:0007155	cell adhesion	Biological Process	21	7.4205	6.51E-07	COL6A4A, DCHS1A, SPON2B, TENM3, ITGB4, ZMP:0000000650, HAPLN2, POSTNA, COMP, PCDH7B, MAG, CCN2A, CCN4B, COL5A1, NID2B, CDH2, LICAMA, ITGA11A, THBS2A, VCANA, CLDN7A	231	437	18397	3.82712711	4.32E-04	2.16E-04	2.14E-04	
GO:0005615	extracellular space	Cellular	35	12.367	1.12E-06	COL16A1, COL5A2A, SOSTDC1A, SEMA3D, MSTNB, COL11A1B, C1QTNF6A, FGF18A, WFDC1, LOXL4, PCSK6, GLDN, LOXA, POSTNA, C4, BMPER, NID2B, CTSS2.1, PCOLCEB, COL1A1A, COL1A1B, ENPP2, CILP2, PRG4B, OLFML2BB, SCARA3, COL10A1A, BGNA, COL27A1B, SFRP2, LFT1, COL5A1, SFRP5, ANGPTL1A, COL14A1A	238	1102	18868	2.51788193	1.48E-04	4.93E-05	4.71E-05	
GO:0043209	myelin sheath	Cellular	5	1.7668	4.80E-06	MBPA, MPZ, HSP90AA1.1, ZWI, PLP1A	238	10	18868	39.6386555	6.34E-04	1.59E-04	1.51E-04	
GO:0005581	collagen trimer	Cellular	8	2.8269	8.84E-06	COL6A4A, COL5A1, COL1A1A, COL10A1A, COL11A1B, COL1A1B, COL14A1A, COL27A1B	238	59	18868	10.7494659	0.00117	2.33E-04	2.23E-04	
GO:0030574	collagen catabolic process	Biological Process	6	2.1201	1.55E-05	MMP11A, MMP2, MMP13B, MMP16B, MMP17B, MMP17A	231	26	18397	18.3786214	0.01026	0.003131	0.0031	
GO:0007422	peripheral nervous system development	Biological Process	6	2.1201	1.89E-05	FOXD3, CDH2, SEMA3D, FGF18A, AGRN, SOX10	231	27	18397	17.6979317	0.01244	0.003131	0.0031	
GO:0005201	extracellular matrix structural constituent	Molecular Function	9	3.1802	2.47E-05	SCARA3, COL16A1, COL5A2A, COL5A1, COL1A1A, COL10A1A, COL11A1B, COL1A1B, COL27A1B	222	93	17340	7.55884917	0.00671	0.00361	0.00358	

GO:0004222	metalloendopeptidase activity	Molecular Function	10	3.5336	2.64E-05	MMP11A, MMP2, MMP13B, ADAMTS17, MMP16B, MMP17B, MMP17A, PAPLN, ADAMTS6, TRABD2B	222	122	17340	6.40230394	0.00719	0.00361	0.00358
GO:0007420	brain development	Biological Process	13	4.5936	4.74E-05	ADGRL2A, MDKB, POU3F1, CORO1A, SCGN, CDH2, AATKA, CNTN2, AATKB, SLC2A3A, CNTN4, CXCR4A, API1M2	231	238	18397	4.35012187	0.03101	0.0063	0.00624
GO:0030424	axon	Cellular Component	11	3.8869	9.15E-05	NTRK1, ADGRL2A, C4, TENM3, ZGC:113337, LICAMA, DNMM1B, MXA, CNTN2, KIF1B, CNTN4	238	179	18868	4.87179006	0.01201	0.002014	0.00192
GO:0007275	multicellular organism development	Biological Process	22	7.7739	1.20E-04	TBX1, IPPK, NOTCH2, NTRK1, SEMA3D, FYNA, SEBOX, COL27A1B, FIBINB, FBLN1, POU3F1, FSTL5, SFRP2, LFT1, PRRX1B, CDH2, LICAMA, MMP11A, SFRP5, SHOX, AGRN, FZD7A	231	675	18397	2.59569665	0.07682	0.01332	0.0132
GO:0030154	cell differentiation	Biological Process	18	6.3604	1.42E-04	SEMA5A, NOTCH2, NTRK1, FOXD3, TENM3, SEMA3D, FYNA, SEBOX, FGF18A, NR2F5, POU3F1, FSTL5, SFRP2, CDH2, LICAMA, SFRP5, AGRN, JAK3	231	487	18397	2.9435985	0.0899	0.013457	0.01334
GO:0014032	neural crest cell development	Biological Process	6	2.1201	3.56E-04	TBX1, FOXD3, SOX8A, SEMA3D, VWA1, SOX10	231	49	18397	9.75192155	0.21068	0.029568	0.0293
GO:0008237	metallopeptidase activity	Molecular Function	9	3.1802	5.00E-04	MMP11A, MMP2, MMP13B, ADAMTS17, MMP16B, MMP17B, MMP17A, ADAMTS6, TRABD2B	222	143	17340	4.91589492	0.12755	0.037454	0.03718
GO:0008201	heparin binding	Molecular Function	7	2.4735	5.49E-04	CCN2A, CCN4B, PCOLCEB, COL5A1, CHAD, THBS2A, MDKB	222	80	17340	6.83445946	0.13917	0.037454	0.03718
GO:0001501	skeletal system development	Biological Process	6	2.1201	7.21E-04	COL1A1A, COL1A1B, COL27A1B, MMP16B, VCANA, HAPLN2	231	57	18397	8.3832308	0.38071	0.053223	0.05274
GO:0005886	plasma membrane	Cellular Component	62	21.908	8.60E-04	GABRB3, ADGRL2A, ATP8A2, TENM3, IGF2BP2A, SI:CH211-286C4.6, HSP90AA1.1, GPR37L1B, PLP1A, SLC01D1, CDH2, DNMM1B, SLC1A2B, MPZ, SLC2A3A, PANX3, CA14, CDON, PLXNA1B, DCHS1A, AGTRAP, PLPPR5B, ACTN2B, PCDH7B, MAG, SLC1A3B, MCOLN1A, TNFRSF21, TRABD2B, SI:CH211-210B2.4, NOTCH2, SEMA3D, FPR1, MMP16B, GRIN1B, CACNA1G, DRP2, ATP1B1B, LICAMA, CD9A, ADGRG6, KCNJ10A, FZD7A, GPR157, PLXNA3, NTRK1, GPR17, GPM6AA, MBPA, ZMP:0000000650, ADGRB1A, GPR22A, MXA, CNTN2, CNTN4, CD247, EFR3BB, RGL1, AGRN, CLDN7A, MYO1F, GRIA1A	238	3299	18868	1.48990399	0.1074	0.016224	0.01549
GO:0014069	postsynaptic density	Cellular Component	7	2.4735	0.001069	CDH2, FAM196A, DNMM1B, MXA, ADGRB1A, ARHGEF9B, SHANK3A	238	92	18868	6.03196931	0.13166	0.017637	0.01684
GO:0005509	calcium ion binding	Molecular Function	21	7.4205	0.001271	NOTCH2, DCHS1A, CD248A, FBLN1, VWDE, ACTN2B, COMP, ANXA13L, FSTL5, PCDH7B, SCGN, PRF1.5, NID2B, CDH2, THBS2A, ENPP2, TNNC1A, CABP2B, VCANA, AGRN, EFCC1	222	739	17340	2.21958088	0.29329	0.069383	0.06887
GO:0007399	nervous system development	Biological Process	11	3.8869	0.001478	NTRK1, FSTL5, IGF2BP2A, CDH2, SEMA3D, LICAMA, IGF2BP3, MDKB, POU3F1, NAV1B, CDON	231	256	18397	3.42206101	0.62555	0.098157	0.09727
GO:0001755	neural crest cell migration	Biological Process	7	2.4735	0.002484	SEMA5A, TBX1, FOXD3, CDH2, SEMA3D, MMP17B, SOX10	231	109	18397	5.11453989	0.80827	0.149966	0.14861

GO:0051252	regulation of RNA metabolic process	Biological Process	3	1.0601	0.006555	<i>IGF2BP2A, PCBP4, IGF2BP3</i>	231	10	18397	23.8922078	0.98731	0.362715	0.35944
GO:0017147	Wnt-protein binding	Molecular Function	4	1.4134	0.007715	<i>SFRP2, SFRP5, FZD7A, TRABD2B</i>	222	32	17340	9.76351351	0.87929	0.351036	0.34846
GO:0042552	myelination	Biological Process	4	1.4134	0.009402	<i>MPZ, FYNA, ADGRG6, PIP1A</i>	231	35	18397	9.10179345	0.99811	0.480205	0.47587
GO:0043025	neuronal cell body	Cellular Component	5	1.7668	0.010554	<i>GPM6AA, ZGC:113337, SNCGA, LICAMA, HSP90AA1.1</i>	238	68	18868	5.82921404	0.75354	0.154794	0.14776
GO:0010001	glial cell differentiation	Biological Process	3	1.0601	0.011085	<i>CDH2, CDKN1CA, SOX10</i>	231	13	18397	18.3786214	0.99939	0.525762	0.52101
GO:0007517	muscle organ development	Biological Process	4	1.4134	0.01356	<i>COL6A4A, MSTNB, HSP90AA1.1, LOXA</i>	231	40	18397	7.96406926	0.99988	0.569595	0.56445
GO:0048675	axon extension	Biological Process	4	1.4134	0.014499	<i>SEMA5A, KIF1B, GRNB, PLXNA3</i>	231	41	18397	7.76982367	0.99994	0.569595	0.56445
GO:0048752	semicircular canal morphogenesis	Biological Process	3	1.0601	0.01468	<i>TBX1, CDH2, SOX10</i>	231	15	18397	15.9281385	0.99995	0.569595	0.56445
GO:0098609	cell-cell adhesion	Biological Process	7	2.4735	0.015441	<i>DCHS1A, CDH2, ITGA11A, CNTN2, ZMP:0000000650, CNTN4, CDON</i>	231	160	18397	3.4842803	0.99997	0.569595	0.56445
GO:0050840	extracellular matrix binding	Molecular Function	3	1.0601	0.017252	<i>BMPER, CD248A, ADGRG6</i>	222	16	17340	14.6452703	0.99136	0.672846	0.66792
GO:0048484	enteric nervous system development	Biological Process	4	1.4134	0.018612	<i>FOXD3, SEMA3D, SEBOX, SOX10</i>	231	45	18397	7.07917268	1	0.650431	0.64455
GO:0033339	pectoral fin development	Biological Process	4	1.4134	0.019731	<i>COL11A1A, SHOX, FGF18A, PDLIM7</i>	231	46	18397	6.92527762	1	0.655057	0.64914
GO:0048538	thymus development	Biological Process	3	1.0601	0.023138	<i>TBX1, SEMA3D, MCM2</i>	231	19	18397	12.5748462	1	0.731602	0.72499
GO:0007166	cell surface receptor signaling pathway	Biological Process	6	2.1201	0.02448	<i>ADGRL2A, ADGRG6, ADGRB1A, CD247, FZD7A, GPR157</i>	231	131	18397	3.64766531	1	0.738846	0.73217
GO:0071526	semaphorin-plexin signaling pathway	Biological Process	4	1.4134	0.025871	<i>SEMA5A, SEMA3D, PLXNA3, PLXNA1B</i>	231	51	18397	6.24632884	1	0.746896	0.74015
GO:0035567	non-canonical Wnt signaling pathway	Biological Process	3	1.0601	0.027956	<i>SFRP2, SFRP5, FZD7A</i>	231	21	18397	11.3772418	1	0.773442	0.76645
GO:0008233	peptidase activity	Molecular Function	12	4.2403	0.028341	<i>CTSS2.1, MMP11A, MMP2, MMP13B, ADAMTS17, PSMB8A, MMP16B, MMP17B, MMP17A, PCSK6, ADAMTS6, TRABD2B</i>	222	446	17340	2.10156345	0.99961	0.967128	0.96004
GO:0007492	endoderm development	Biological Process	3	1.0601	0.030502	<i>CCN2A, FRAS1, SEBOX</i>	231	22	18397	10.8600945	1	0.810136	0.80282
GO:0016055	Wnt signaling pathway	Biological Process	6	2.1201	0.035704	<i>SFRP2, SOSTDC1A, SFRP5, CCDC88C, FZD7A, TRABD2B</i>	231	145	18397	3.29547694	1	0.910938	0.90271

GO:0060272	embryonic skeletal joint morphogenesis	Biological Process	2	0.7067	0.037041	SEMA3D, PLXNA3	0.910938	0.90271
GO:0005588	collagen type V trimer	Cellular Component	2	0.7067	0.037213	COL5A2A, COL5A1	0.491217	0.46889
GO:0035118	embryonic pectoral fin morphogenesis	Biological Process	3	1.0601	0.047543	CDH2, FGF18A, FIBINB	1	0.99246
GO:0050931	pigment cell differentiation	Biological Process	2	0.7067	0.049082	FOXD3, SOX10	1	0.99246
GO:0043583	ear development	Biological Process	2	0.7067	0.049082	TBX1, ADGRG6	1	0.99246

## 6 General discussion

The electric fish have long been considered as an excellent model to study convergent evolution, radiative evolution, and electrophysiology. There are more than 180 described species in the African weakly electric fish mormyrids. Such a successful radiative speciation in mormyrids is associated with a highly divergent electric organ discharge (EOD). The major objective of this thesis was to understand the genetic basis of the EOD divergence among mormyrid fish from the gene expression to whole genome level. The provided novel reference genome is also a valuable resource for future work investigating the evolution of this taxon on many levels.

### 6.1 Genomics of *Campylomormyrus*

A high quality whole genome was produced from the species *Campylomormyrus compressirostris*. This is the second well-annotated genome in mormyrids, the fourth genome of an electric fish, and the third well-annotated genome in Osteoglossomorpha. Compared to the available genomes from *Paramormyrops kingsleyae*, *Electrophorus electricus* and *Scleropages formosus*, the contig length from the genome of *C. compressirostris* was significantly improved. The major finding from the genome was the substantially higher number of the predicted protein-coding genes, which counted for 34,492, compared to other annotated Osteoglossomorpha genomes (~23k protein-coding genes). This profound difference could relate to our improved assembly, but could also reflect true differences in *Campylomormyrus*, relative to *Paramormyrops*.

I curated 16 genes from *Kv1* subfamily. The most interesting finding was a further tandem duplication of the *KCNA7a* gene. This duplication was also found in *P. kingsleyae* after re-analyzing its genome, but not in *S. formosus* (a non-electric fish close to mormyrids), nor in other electric fish lineages beyond mormyrids, indicating this tandem duplication might be lineage-specific (perhaps shared among mormyrids). *Kv1* genes encode for voltage-gated potassium

channels, which are considered to be regulate EOD signals in mormyrids. The abundant gene number in this subfamily likely resulted from two rounds of whole genome duplication and an additional teleost-specific whole genome duplication. Those duplicated gene copies might have undergone neofunctionalization or subfunctionalization leading to the retention of those copies. This can be supported by the up-regulation of 13 *Kv1* genes in the electric organ (EO) of *C. tshokwe* compared with skeletal muscle (SM). Two tandem duplicated *KCNA7a* genes were both up-regulated in the EO of *C. compressirostris*, *C. tshokwe*, *C. rhynchophorus* and hybrids *C. compressirostris* x *C. rhynchophorus*, *C. compressirostris* x *C. tshokwe*. *KCNA7a\_2*, which was inferred to be under positive selection in the transmembrane helices 3-4 linkers among species *Brienomyrus* and *Gymnarchus*, exhibited even higher expression than the other copy. This indicates that both *KCNA7a* copies are functional in the EO of *Campylomormyrus* (and perhaps across mormyrids), which resulted in the retention of both copies.

## **6.2 What makes the EO?**

Most electric fish possess a specific myogenic EO derived from SM fibers, indicating a convergent evolution in those independently evolved lineages. Therefore, a transcriptional comparison between EO and SM helped investigate the transitions from SM to EO at the gene expression level during development.

Several EO-specific candidate genes showed convergent expression pattern in the studied *Campylomormyrus* and other electric fish lineages Mormyroidea (beyond *Campylomormyrus*), Siluriformes and Gymnotiformes, e.g. *SCN4aa*, *SIX2a*, *HEY1*, *KCNA7a*, *NDRG3*, *MEF2a* and several isoforms of sodium/potassium ATPase  $\alpha$  and  $\beta$  subunits (Gallant et al. 2012; Gallant et al. 2014; Traeger et al. 2017; Losilla et al. 2020). These consistently up-regulated genes in the EO of electric fish lineages indicate that EOs differ from SM in part due to the specific expression of

sodium and potassium channels, transduction signals, and skeletal muscle-specific transcription factors in general.

In addition to the convergently expressed genes in the EO in general, *Campylomormyrus* also exhibited specific gene expression patterns. This is evidenced by the abundantly up-regulated F-actin dynamics, unconventional myosin and *MEF2* genes. Those motoric/sarcomeric genes were quite different relative to another mormyrid, *Brienomyrus brachyistius*, where, e.g. troponin I isoforms, myosin heavy chain and tropomyosin were specifically up-regulated in the EO (Gallant et al. 2012). It suggested that the organization of the F-actin system in electrocytes might vary across mormyrid genera.

One transcription factor (*KLF5* - Krüppel-Like Factor 5) was highly overexpressed in the EO of *Campylomormyrus*. Krüppel is involved in the regulation of potassium channel expression in *Drosophila* when there is no expression from *KCND* genes (Parrish et al. 2014). We did not observe the expression of *KCND* genes in the EO of *Campylomormyrus* as well. In addition, several potassium channel genes were up- or down-regulated in the EO. We therefore suppose that the Krüppel transcription factor (*KLF5* here) possibly contributes to the regulation of potassium channels in the EO of *Campylomormyrus*, thereby enabling the formation of a special SM-derived EO.

### **6.3 EOD duration diversification among *Campylomormyrus* species**

The EOD is usually species-specific in mormyrids with regard to shape and waveform. The diverged EOD serves as a pre-zygotic isolation mechanism and hence contributes to the species radiation in mormyrids. The EOD is assumed to be regulated by different sodium and potassium currents across the plasma membrane. We clearly see gene expression patterns related to the EOD



duration, consistently observed in both adult and juvenile EO samples (Article II, Fig. 3a; Article III, Fig. 2a).

We identified several candidate genes that showed an increasing or decreasing expression pattern relative to the EOD duration, including three potassium channel genes (*KCNJ2*, *KCNK6* and *KCNQ5*) and a transcription factor *KLF5* (Article II, Table 2). Both *KCNK6* and *KCNQ5* encode for the voltage-gated potassium channel genes, which generate outward potassium currents (Schroeder et al. 2000; Chai et al. 2017). *KCNK5*, which is a paralog of *KCNK6*, also shows higher expression in *Paramormyrops* species with short EOD (Losilla et al. 2020). Their lower expression in species with elongated EOD might result in decreased outward potassium currents and eventually prolongate the repolarization of the EOD. *KCNJ2* encodes for an inward rectifying potassium channel. It stabilizes the resting membrane potential, shaping the initial depolarization and final repolarization of cardiomyocytes' action potential (Dhamoon and Jalife 2005; Hibino et al. 2010). The expression of this gene is higher in elongated EOD species, along with the transcription factor *KLF5*.

We amplified *KCNJ2* genes in 10 *Campylomormyrus* species using different pairs of primers designed from the genome (this part was performed by a bachelor student; Cheri, Cheng & Tiedemann, unpubl. results). Two non-synonymous substitutions were identified in *C. rhynchophorus* and *C. numenius*. The substitution at site 198 was predicted change the function of the encoded protein. This suggests that the *KCNJ2* gene has a dual impact on the modulation of EOD duration, resulting from both altered gene expression and protein function. Those two species possess very elongated EOD (over 40 ms). They have a special papillae structure at the anterior face of the electrocyte. We do not know the relationship between papillae and inwardly rectifier potassium channels (encoded by the *KCNJ2* gene). It is possible that this special folding and

invagination on the membrane might increase membrane capacity and increase ion currents to elongate the EOD. Further exploration on electrophysiology and histology is essential to prove the function of *KCNJ2* on EOD modulation.

#### **6.4 EOD development during ontogeny**

The mormyrid fish possess a larval EO in the deep lateral muscle (Nguyen et al. 2020). The EOD produced by the larval organ is biphasic and short (around 0.4 ms) duration in several studied *Campylomormyrus* species. Throughout the ontogeny of *C. compressirostris*, the EOD shows no change in shape and duration. However, in *C. rhynchophorus* and *C. tshokwe*, the EOD exhibits continuous alteration throughout several stages from larval EOD until the adult EOD.

We observed several potassium channel genes that were differentially expressed throughout the ontogeny of *C. compressirostris*, *C. rhynchophorus* and their cross-species hybrid (also exhibiting EOD change during ontogeny). *KCNJ2* also showed higher expression in the adult *C. rhynchophorus* (40 ms duration EOD) and the hybrid *C. compressirostirs x C. rhynchophorus* (4 ms duration EOD), compared to their respective juveniles, which have 5 ms and 0.4 ms duration EODs respectively. This further suggests the *KCNJ2* gene to contribute to the elongation of EOD in *Campylomormyrus*. In addition, several voltage-gated potassium channel genes exhibited a decreasing expression pattern throughout ontogeny. They may generate lower outward potassium currents and consequently elongated EOD during development.

Although *C. compressirostris* does not show any EOD alteration during ontogeny, we still observed some potassium channel genes (e.g. *KCNA6a*) to be down-regulated in the adult. One possible explanation is that the membrane capacity may have enlarged following the growth of the electrocytes growth. The down-regulated voltage-gated potassium channel gene might compensate

for this developmental change and result in the EOD consistency during ontogeny. However, this hypothesis needs further investigation.

### **6.5 EOD diversification in *Campylomormyrus* hybrids**

Hybrids play an important role in understanding the phenotypic evolution, as they contain two different subgenomes from their respective parental species. The *Campylomormyrus* hybrids exhibited interesting EODs compared to their parental species. We explored the gene and allele specific expression of two hybrid cohorts, i.e., crosses among the short duration EOD species *C. compressirostris* and (1) a medium duration EOD species (*C. tshokwe*) resp. (2) an elongated duration EOD species (*C. rhynchophorus*). Short duration EOD (e.g. in adult *C. compressirostris* and *C. tamandua*) is considered as an ancestral phenotype (plesiomorphic), while the long duration EOD (e.g. medium duration EOD in *C. tshokwe* and *C. alces*; elongated duration EOD in *C. rhynchophorus* and *C. numenius*) as a derived phenotype (apomorphic).

Both hybrids exhibit intermediate EODs in the adult stage compared to their parental species. However, the shape and duration of hybrids' EODs all more closely resemble their parental species with elongated EOD instead of *C. compressirostris*, i.e., they had a tendency towards the apomorphic EOD phenotype.

In the allele specific analyses on the EO and SM of two adult hybrids, we observed a general dominance of the alleles from the *C. compressirostris* (Article II, Fig. 5a). Such dominance occurred in the EO of juvenile hybrids *C. compressirostris* x *C. rhynchophorus* as well (Article III, Fig. 3a). We inferred that the subgenome from the parental species with a plesiomorphic EOD appeared dominantly expressed in the hybrids.

We identified several genes that showed allele expression imbalance in the adult hybrids, with only one gene (*KCNJ2*) exhibiting dominant expression of the allele from *C. rhynchophorus* instead of *C. compressirostris* (Article II, Fig. 5b). In addition, the expression of the *C. rhynchophorus* allele increased during ontogeny of the hybrid *C. compressirostris* x *C. rhynchophorus* (Article III, Fig. 3c). This gene showed a higher expression in the purebred species with an elongated EOD, suggesting its expression might be under cis-regulation in both purebred species and hybrids. The allele specific expression results support the potential function of the *KCNJ2* gene in regulating EOD duration of *Campylomormyrus*.

This is the first time to explore the allele specific expression in the electric fish hybrids and ontogeny. Although, the asymmetric expression of two parental alleles affect the phenotype was studied in other species. The F1 hybrids between California tiger salamander and barred tiger salamander show that the proportion of genes with allelic imbalance might relate to their thermal tolerance (Cooper and Shaffer 2021). The Amazon molly also exhibits allelic bias favoring the maternal ancestor allele in ovarian tissue (Zhu et al. 2017). Those studies suggest that the allele specific expression (allele imbalanced expression) has an impact on the phenotype in hybrids.

Therefore, the allele imbalanced expression of *KCNJ2* gene probably affects the EOD resemblance to *C. rhynchophorus* in the hybrid *C. compressirostris* x *C. rhynchophorus* under the scenario of overall dominance of *C. compressirostris* alleles, and also might affect the EOD development of this hybrid during ontogeny by increasing the expression of the allele from *C. rhynchophorus*.

## 7 Conclusion

My PhD study produced a high quality genome in *Campylomormyrus*. This is the fourth well-annotated genome in electric fish and third in Osteoglossomorpha. It will be a valuable resource for understanding the evolution of electric fish as well as teleosts.

This study further investigated the EOD diversification across species and ontogeny of *Campylomormyrus* species/hybrids from a gene expression perspective. Several candidate genes potentially involved in the EOD waveform were identified, e.g. *KCNJ2*, *KLF5*, *KCNK6*, *KCNQ5*. *KCNJ2* is also involved in the EOD development during ontogeny. In addition, I identified one transcription factor *KLF5* that might regulate different potassium channels expression in *Campylomormyrus*.

The alleles from *C. compressirostris* exhibited near global expression dominance in the hybrids. A single exception was *KCNJ2*, which showed an opposite expression dominance of the allele from *C. rhynchophorus* in their hybrids. This further supported our hypothesis that *KCNJ2* is a powerful candidate in regulating EOD duration.

In summary, the major finding figured out the candidate genes that affect EOD duration diversification in the weakly electric fish. It contributed to our understanding of the adaptive radiation in Mormyroidae (even Osteoglossomorpha) under a possibly sympatric condition.

Further exploration is necessary to prove our hypotheses on the inferred EOD duration-specific candidate genes. Gene knockdown on *KCNJ2* in different *Campylomormyrus* will be useful to verify the function on EOD duration regulation. In addition, gene editing in *C. rhynchophorus*/*C. numenius* and hybrids will help us to understand if the non-synonymous mutation will alter the protein function in very elongated EOD species, and perhaps will change the EOD phenotype of

hybrids. Different sets of hybrids crossing species with plesiomorphic and apomorphic EOD features might help to understand if the subgenome from parental species with plesiomorphic traits are generally more dominant in the hybrid offspring. More detailed gene expression during the ontogeny of *C. rhynchophorus* and *C. numenius* might help us to understand the EOD development.

## 8 Reference

- Arnegard ME, Bogdanowicz SM, Hopkins CD. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. *Evolution (N Y)* 59:324–343.
- Arnegard ME, McIntyre PB, Harmon LJ, Zelditch ML, Crampton WGR, Davis JK, Sullivan JP, Lavoué S, Hopkins CD. 2010. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *American Naturalist* 176:335–356.
- Bass AH. 1986. Electric Organs Revisited: Evolution of a Vertebrate Communication and Orientation Organ.
- Bennett MVL. 1970. Comparative physiology: electric organs. *Annu Rev Physiol* 32:471–528.
- Bennett MVL. 1971. Electric organs. In: *Fish Physiology*. Vol. 5. p. 347–491.
- Carlson BA, Hasan SM, Hollmann M, Miller DB, Harmon LJ, Arnegard ME. 2011. Brain evolution triggers increased diversification of electric fishes. *Science (1979)* 332:583–586.
- Carlson BA, Sisneros JA, Popper AN, Fay RR. 2019. Electroreception : fundamental insights from comparative approaches. Available from:  
[https://www.google.com/books/edition/\\_/zCy-DwAAQBAJ?hl=en](https://www.google.com/books/edition/_/zCy-DwAAQBAJ?hl=en)
- Chai S, Wan X, Nassal DM, Liu H, Moravec CS, Ramirez-Navarro A, Deschênes I. 2017. Contribution of two-pore K<sup>+</sup> channels to cardiac ventricular action potential revealed using human iPSC-derived cardiomyocytes. *Am J Physiol Heart Circ Physiol* 312:H1144–H1153.
- Crampton WGR. 2019. Electroreception, electrogenesis and electric signal evolution. *J. Fish Biol.* 95:92–134.
- Dhamoon AS, Jalife J. 2005. The inward rectifier current (IK1) controls cardiac excitability and is involved in arrhythmogenesis. *Heart Rhythm* 2:316–324.
- Feulner PGD, Kirschbaum F, Tiedemann R. 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.

- Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R. 2009a. Electrifying love: Electric fish use species-specific discharge for mate recognition. *Biol Lett* 5:225–228.
- Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R. 2009b. Magic trait electric organ discharge (EOD): Dual function of electric signals promotes speciation in African weakly electric fish. *Commun. Integr. Biol.* 2:329–331.
- Gallant JR, Hopkins CD, Deitcher DL. 2012. Differential expression of genes and proteins between electric organ and skeletal muscle in the mormyrid electric fish *Brienomyrus brachyistius*. *Journal of Experimental Biology* 215:2479–2494.
- Gallant JR, Traeger LL, Volkening JD, Moffett H, Chen P, Novina CD, Jr GNP, Anand R, Wells GB, Pinch M, et al. 2014. Genomics basis for the convergent evolution of electric organs. *Science (1979)* 344:1522–1525.
- Glasauer SMK, Neuhauss SCF. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics* 289:1045–1060.
- Gundappa MK, To TH, Grønvold L, Martin SAM, Lien S, Geist J, Hazlerigg D, Sandve SR, Macqueen DJ. 2022. Genome-Wide Reconstruction of Rediploidization Following Autopolyploidization across One Hundred Million Years of Salmonid Evolution. *Mol Biol Evol* 39:msab310.
- Hibino H, Inanobe A, Furutani K, Murakami S, Findlay I, Kurachi Y. 2010. Inwardly rectifying potassium channels: Their structure, function, and physiological roles. *Physiol Rev* 90:291–366.
- Hooper SD, Berg OG. 2003. On the nature of gene innovation: Duplication patterns in microbial genomes. *Mol Biol Evol* 20:945–954.
- Jatllon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Maucell E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.
- Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719.



- Kirschbaum F, Nguyen L, Baumgartner S, Chi HWL, Wolfart R, Elarbani K, Eppenstein H, Korniienko Y, Guido-Böhm L, Mamonekene V, et al. 2016. Intra-genus (Campylomormyrus) and intergenus hybrids in mormyrid fish: Physiological and histological investigations of the electric organ ontogeny. *J Physiol Paris* 110:281–301.
- Korniienko Y, Tiedemann R, Vater M, Kirschbaum F. 2021. Ontogeny of the electric organ discharge and of the papillae of the electrocytes in the weakly electric fish *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *Journal of Comparative Neurology* 529:1052–1065.
- Lamanna F, Kirschbaum F, Ernst ARR, Feulner PGD, Mamonekene V, Paul C, Tiedemann R. 2016. Species delimitation and phylogenetic relationships in a genus of African weakly-electric fishes (Osteoglossiformes, Mormyridae, Campylomormyrus). *Mol Phylogenet Evol* 101:8–18.
- Lamanna F, Kirschbaum F, Waurick I, Dieterich C, Tiedemann R. 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* [Internet] 16:1–17. Available from: <http://dx.doi.org/10.1186/s12864-015-1858-9>
- Losilla M, Luecke DM, Gallant JR. 2020. The transcriptional correlates of divergent electric organ discharges in *Paramormyrops* electric fish. *BMC Evol Biol* 20:1–19.
- Nagel R, Kirschbaum F, Engelmann J, Hofmann V, Pawelzik F, Tiedemann R. 2018. Male-mediated species recognition among african weakly electric fishes. *R Soc Open Sci* 5:4–11.
- Nagel R, Kirschbaum F, Hofmann V, Engelmann J, Tiedemann R. 2018. Electric pulse characteristics can enable species recognition in African weakly electric fish species. *Sci Rep* 8:1–12.
- Nagel R, Kirschbaum F, Tiedemann R. 2017. Electric organ discharge diversification in mormyrid weakly electric fish is associated with differential expression of voltage-gated ion channel genes. *Journal of Comparative Physiology A* 203:183–195.

- Nguyen L, Mamonekene V, Vater M, Bartsch P, Tiedemann R, Kirschbaum F. 2020. Ontogeny of electric organ and electric organ discharge in *Campylomormyrus rhynchophorus* (Teleostei: Mormyridae). *Journal of Comparative Physiology A* 206:453–466.
- Nguyen L, Paul C, Mamonekene V, Bartsch P, Tiedemann R, Kirschbaum F. 2017. Reproduction and development in some species of the weakly electric genus *Campylomormyrus* (Mormyridae, Teleostei). *Environ Biol Fishes* 100:49–68.
- Parrish JZ, Kim CC, Tang L, Bergquist S, Wang T, DeRisi JL, Jan LY, Jan YN, Davis GW. 2014. Krüppel Mediates the Selective Rebalancing of Ion Channel Expression. *Neuron* 82:537–544.
- Paul C, Mamonekene V, Vater M, Feulner PGD, Engelmann J, Tiedemann R, Kirschbaum F. 2015. Comparative histology of the adult electric organ among four species of the genus *Campylomormyrus* (Teleostei: Mormyridae). *Journal of Comparative Physiology A* 201:357–374.
- Schroeder BC, Hechenberger M, Weinreich F, Kubisch C, Jentsch TJ. 2000. KCNQ5, a Novel Potassium Channel Broadly Expressed in Brain, Mediates M-type Currents. *Journal of Biological Chemistry* 275:24089–24095.
- Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet* 14:751–764.
- Swapna I, Ghezzi A, York JM, Markham MR, Halling DB, Lu Y, Gallant JR, Zakon HH. 2018. Electrostatic Tuning of a Potassium Channel in Electric Fish. *Current Biology* 28:2094–2102.e5.
- Thompson A, Vo D, Comfort C, Zakon HH. 2014. Expression evolution facilitated the convergent neofunctionalization of a sodium channel gene. *Mol Biol Evol* 31:1941–1955.
- Tiedemann R, Feulner PGD, Kirschbaum F. 2010. Electric Organ Discharge Divergence Promotes Ecological Speciation in Sympatrically Occurring African Weakly Electric Fish (*Campylomormyrus*). In: *Evolution in Action: Case studies in Adaptive Radiation, Speciation and the Origin of Biodiversity*. p. 307–321.
- Traeger LL, Sabat G, Barrett-Wilt GA, Wells GB, Sussman MR. 2017. A tail of two voltages: Proteomic comparison of the three electric organs of the electric eel. *Sci Adv* 3:e1700523.

- Volff JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb)* 94:280–294.
- Wang Y, Yang L. 2021. Genomic Evidence for Convergent Molecular Adaptation in Electric Fishes. *Genome Biol Evol* 13:1–11.
- Zhu F, Schlupp I, Tiedemann R. 2017. Allele-specific expression at the androgen receptor alpha gene in a hybrid unisexual fish, the Amazon molly (*Poecilia formosa*). *PLoS One* 12:1–14.

## **9 Acknowledgements**

I would firstly thank my first supervisor Prof. Dr. Ralph Tiedemann for offering me the opportunity to work on this interesting and challenging topic in the Evolutionary Biology/Systematic Zoology group. I appreciate his professional supervision and support during my five years here.

I also want to thank Prof. Dr. Frank Kirschbaum for sharing his knowledge of weakly electric fish. He helped a lot in fish breeding experiment side by side. Great appreciation also give to apl. Prof. Dr. Otto Baumann for sharing his electrophysiology knowledge.

Many thanks go to apl. Prof. Dr. Alice B. Dennis for her supervision on experiment design, bioinformatics and analysis guidance. Her excellent knowledge on evolutionary biology, bioinformatics helped me a lot. I also would like to thank Dr. Stefanie Hartmann and Dr. Christian Kappel for their help on bioinformatics and data analyses.

Special thanks go to all the colleagues at the group Evolutionary Biology/Systematic Zoology for the relaxed atmosphere. I like to thank Dr. Linh Nguyen for taking care of the breeding fish as well as sharing his knowledge of weakly electric fish. I would like to thank our technicians Anja Ernst, Katja Havenstein and Tonio Pieterek for their help in the lab, to Dr. Marisol Domínguez for her suggestion to my project. To Dr. Julia Cantiz, Dr. Rahma Amen and Andrew Sinnott, specially thank them for building a nice office together.

Finally, I would like to thank my family and Xuan for having backed on me during my PhD.