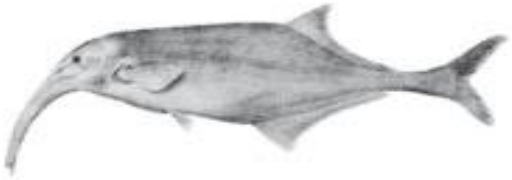
Institut für Biochemie und Biologie
Evolutionsbiologie/Spezielle Zoologie

Adaptive radiation and speciation in African weakly electric fish: a phylogenetic and transcriptomic perspective.

(Genus *Campylomormyrus,* Mormyridae, Teleostei)



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
Francesco Lamanna

Potsdam, März 2015

Gutachter:

• **1. Gutachter:** Prof. Dr. Ralph Tiedemann (Universität Potsdam)

• **2. Gutachter:** Prof. Dr. Walter Salzburger (Universität Basel)

• **3. Gutachter:** PD Dr. Frieder Mayer (Leibniz-Institut für Evolutions- und Biodiversitätsforschung am Museum für Naturkunde, Berlin)

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-80097
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-80097

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.

Potsdam, den

Francesco Lamanna

Table of contents

1 Abstract

The rise of evolutionary novelties is one of the major drivers of evolutionary diversification. African weakly-electric fishes (Teleostei, Mormyridae) have undergone an outstanding adaptive radiation, putatively owing to their ability to communicate through species-specific Electric Organ Discharges (EODs) produced by a novel, muscle-derived electric organ. Indeed, such EODs might have acted as effective pre-zygotic isolation mechanisms, hence favoring ecological speciation in this group of fishes. Despite the evolutionary importance of this organ, genetic investigations regarding its origin and function have remained limited.

The ultimate aim of this study is to better understand the genetic basis of EOD production by exploring the transcriptomic profiles of the electric organ and of its ancestral counterpart, the skeletal muscle, in the genus *Campylomormyrus*. After having established a set of reference transcriptomes using "Next-Generation Sequencing" (NGS) technologies, I performed *in silico* analyses of differential expression, in order to identify sets of genes that might be responsible for the functional differences observed between these two kinds of tissues. The results of such analyses indicate that: i) the loss of contractile activity and the decoupling of the excitation-contraction processes are reflected by the down-regulation of the corresponding genes in the electric organ; ii) the metabolic activity of the electric organ might be specialized towards the production and turn-over of membrane structures; iii) several ion channels are highly expressed in the electric organ in order to increase excitability, and iv) several myogenic factors might be down-regulated by transcription repressors in the EO.

A secondary task of this study is to improve the genus level phylogeny of *Campylomormyrus* by applying new methods of inference based on the multispecies coalescent model, in order to reduce the conflict among gene trees and to reconstruct a phylogenetic tree as closest as possible to the actual species-tree. By using 1 mitochondrial and 4 nuclear markers, I was able to resolve the phylogenetic relationships among most of the currently described *Campylomormyrus* species. Additionally, I applied several coalescent-based species delimitation methods, in order to test the hypothesis that putatively cryptic species, which are distinguishable only from their EOD, belong to independently evolving lineages. The results of this analysis were additionally validated by investigating patterns of diversification at 16 microsatellite loci. The results suggest the presence of a new, yet undescribed species of *Campylomormyrus*.

2 Zusammenfassung

Das übergreifende Ziel dieser Arbeit ist das bessere Verständnis der Bedeutung der schwachen Elektrizität für die adaptive radiation der Mormyriden Afrikas. Das gewählte Modell-Taxon, die Mormyriden-Gattung *Campylomormyrus*, zeigt eine große Vielfalt an elektrischen Entladungsformen. Diese Entladungsformen sind artspezifisch. Die genetische Basis dieses Merkmales ist allerdings noch unbekannt. In dieser Arbeit habe ich transkriptomische Untersuchungen vom elektrischen Organ und Skelettmuskel durchgeführt. Die Ergebnisse dieser Analysen zeigen, dass die phenotypischen Unterschiede zwischen dem elektrischen Organ und dem Skelettmusckel in den jeweiligen transkriptomen gespiegelt sind.

Ich habe auch einen phylogenetischen Stammbaum für die Gattung *Campylomormyrus* hergestellt, durch die Anwendung von „Multispecies Coalescent Models"-basierten Methoden. Außerdem, durch die Anwendung von Mikrosatellitdaten, die als unabhängiger Beweis dienten, konnte ich zeigen, dass die identifizierten phylogenetischen Gruppen reproduktiv isolierte biologische Arten sind. Auf diese Weise konnte ich ein neuen, noch unbeschriebenen Art nachweisen.

3 Introduction.

3.1 Systematic placement of the African weakly-electric fish genus *Campylomormyrus*.

Mormyrids are a family of freshwater weakly-electric fishes endemic to African riverine and, partially, lacustrine systems. They belong to the superorder Osteoglossomorpha, considered one of the phylogenetically basal groups of Teleostei (Near *et al.* 2012). With more than 180 described species in 20 genera, mormyrid fishes account for almost 90% of the overall diversity within osteoglossomorphs (Lavoué & Sullivan 2004). Together with the monospecific family Gymnarchidae, they constitute the superfamily Mormyroidea (Fig. 1); the most remarkable synapomorphic (derived) characters supporting the monophyly of this taxon are definitely the presence of electric organs, matched electroreceptors and a substantially enlarged cerebellum (Taverne 1972). The monophyletic status of Mormyroidea has been further confirmed by several molecular phylogenetic studies (Alves-Gomes & Hopkins 1997; Sullivan *et al.* 2000). Within the family Mormyridae two subfamilies are recognized (Fig. 1): Petrocephalinae, formed by a single genus (*Petrocephalus*) composed of 42 species and Mormyrinae constituted by all the remaining genera and species.

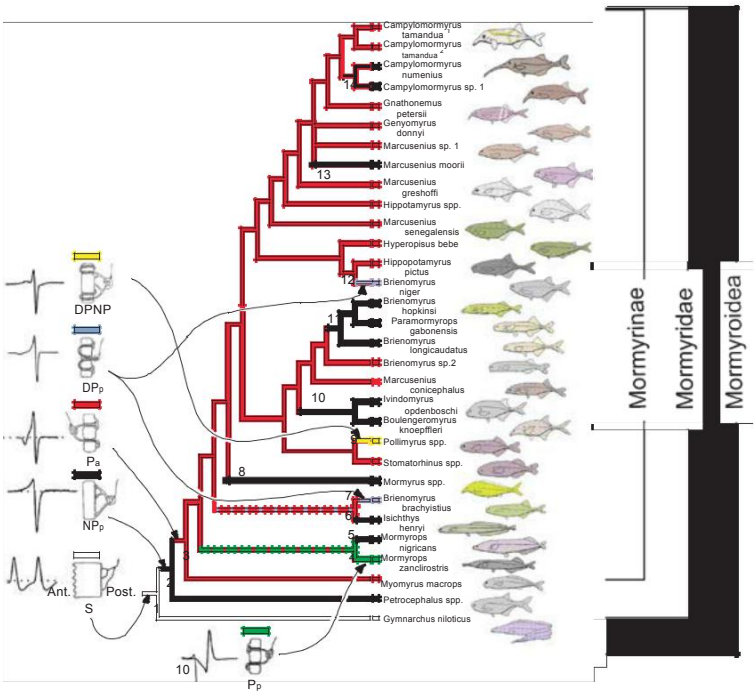


Figure 1 Phylogenetic tree of Mormyroidea. On the left side the different electrocyte morphologies and Electric Organ Discharge (EOD) types are represented, as well as their position on the phylogeny. The abbreviations below each electrocyte indicate its morphology and innervation pattern: S=stalkless; NP_p=non penetrating stalks posteriorly innervated; P_a =penetrating stalks anteriorly innervated; DP_p=doubly penetrating stalks posteriorly innervated; DPNP=doubly penetrating and non-penetrating stalks; P_p=penetrating stalks posteriorly innervated. Ant.=anterior side of the electrocyte; Post.=posterior side (Modified from Sullivan *et al.* 2000).

One of the most peculiar genera of the subfamily Mormyrinae is *Campylomormyrus.* This genus

RULJLQDWHG LQ UHODWLYHO\ UHFHQW WLPHV § 0ya; Lavoué *et al.* 2012) and is formed by 15 described species (Poll *et al.* 1982; Feulner *et al.* 2007). All members of the genus *Campylomormyrus* are characterized by the presence of prominent, tubular, elongated snouts (Fig. 2). All species are endemic to the Congo River basin, with the highest peaks of diversity observed in the rapids occurring along the last 300 km upstream of the river mouth. Over the past years, few molecular phylogenies of the genus were produced, which allowed to identify several monophyletic groups corresponding to well defined body morphologies and electric features (Feulner *et al.* 2006, 2007). These studies, however, did not allow to resolve the relationships between few terminal taxa (Fig. 2), associated to short branches, that may have being prone to extensive levels of Incomplete Lineage Sorting (ILS; Pamilo & Nei 1988).

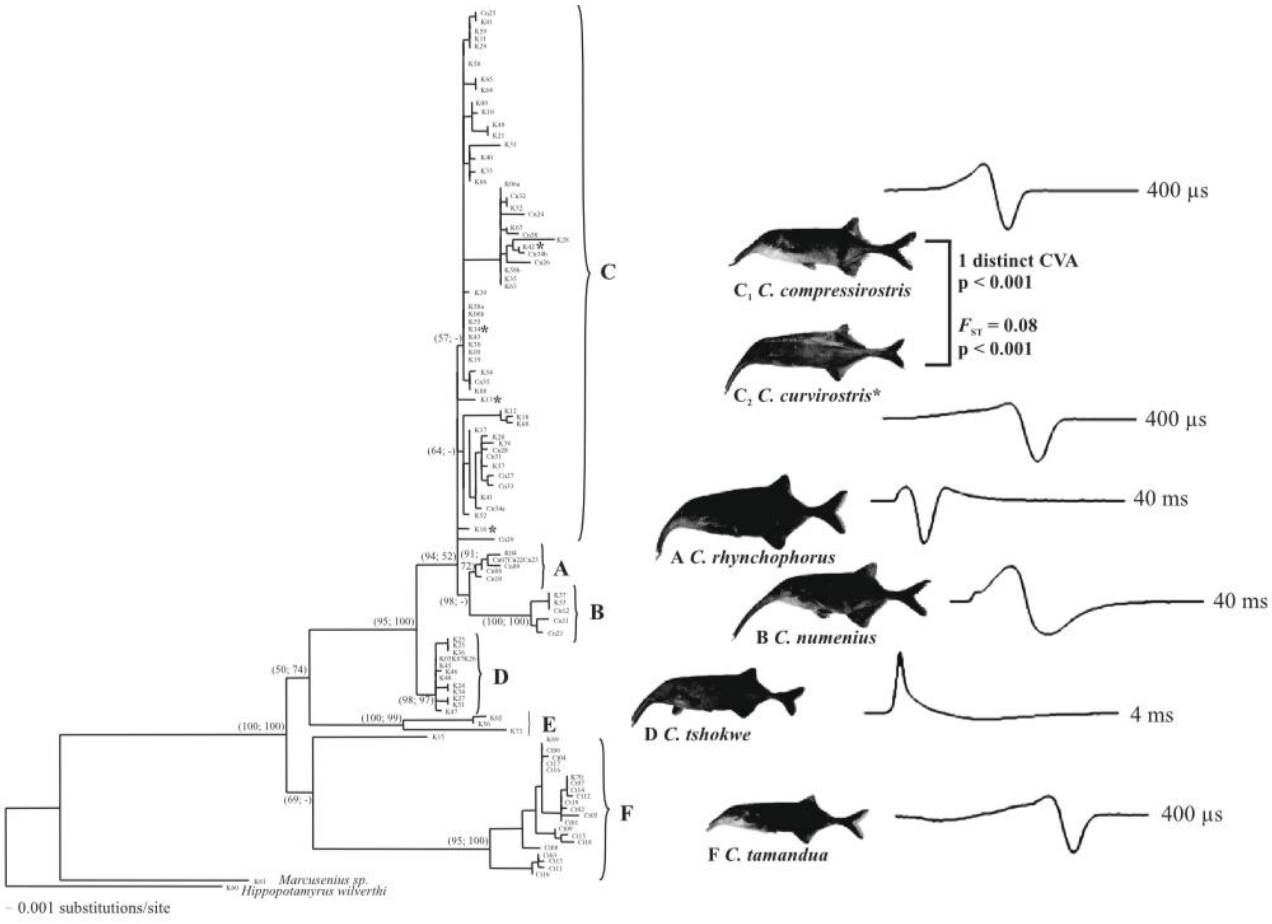


Figure 2 Bayesian phylogeny based on the combined data set of mitochondrial cytochrome b and nuclear S7 genes (Feulner *et al.* 2007). Representative photographs and adult electric organ discharges are shown for identified species within the tree. In case of group C the differentiation between *C. compressirostris* and *C. curvirostris* is not resolved in the phylogenetic tree.

3.2 On Electrogenesis and electroreception.

Bioelectrogenesis (i.e. the ability to produce electric fields by specialized organs) has evolved several times independently in aquatic vertebrates (Bass 1986). In fact, it can be observed in the marine electric rays (Torpediniformes) and skates (Rajiformes), in the African freshwater Mormyridae and Gymnarchidae (Mormyroidea), in the South American knifefishes (Gymnotiformes), in several catfish species (Siluriformes), and in few marine stargazers (Perciformes). In all the above-mentioned groups, electric organs originate from myogenic tissue; the only exception are members of the family Apteronotidae (Gymnotiformes), where the electric organs are formed by modified spinal motor neurons (Kirschbaum 1983). In mormyrid fishes, the cells forming the electric organ are compressed, disk-like, cells commonly called electrocytes (see figs. 1 and 3). They are longitudinally stacked behind each other in order to form columns of cells embedded within tubes of isolating connective tissue. The synchronous activity of each electrocyte defines the output of the electric organ, known as Electric Organ Discharge (EOD, Lissmann 1958). The amount of electrocytes within each electric organ determines the electric potential of an EOD, which can range from few millivolts (weakly-electric fish) to several hundreds of volts (strongly-electric fish). Strongly electric fishes use their powerful EOD mainly for self-defence or predation purposes; well-known examples are the electric ray *Torpedo* spp., the electric catfish *Malapterurus africanus* and the electric eel *Electrophorus electricus*. Weakly-electric fishes, on the other hand, use their electric sense for: i) the localization and discrimination of conductive objects in water (active electroreception, Lissmann & Machin 1958); ii) the recognition of conspecific individuals (Feulner *et al.* 2009a); and iii) in social and reproductive behavior (Bratton & Kramer 1989; Crawford 1991). Weakly-electric fishes can be further subdivided into: wave-discharging type and pulse-discharging type. Wave-discharging fishes continuously produce a wave-like, phase-locked, discharge at a constant frequency, whereas pulse-discharging fishes produce pulse-like EODs at irregular intervals; all mormyrids are pulse-type dischargers.

In all African weakly-electric fishes, the electric organ is located in the caudal peduncle and is formed by four columns of electrocytes, two dorsal and two ventral ones (Fig. 3). Each electrocyte is innervated by electromotoneurons originating in the spinal cord. Electric organs arise in juvenile fishes from several myomeres of the deep lateral muscle; their myogenic origin is confirmed by the presence of disorganized myofibrils within the electrocytes (Szabo 1960; Denizot *et al.* 1982). Electrocyte anatomy is substantially variable among mormyrid fish species and influences EOD shape and duration. The main anatomical differences are observed within the innervation patterns and in the so-called “stalk system”: a system of cell membrane protrusions that can vary dramatically even between closely related species (Moller 1995) (Figs.1 and 3).

Two classes of tubular electroreceptors are used by mormyrid fishes: knollenorgans and mormyromasts. Knollenorgans respond to both the fish’s own and other individuals’ EODs, and are

used for species and sex discrimination. Mormyromasts, conversely, are responsible for active electrolocation, that is, for sensing the transcutaneous current evoked by the fish's own EOD and the alteration of such currents by external objects (Bullock *et al.* 2005).

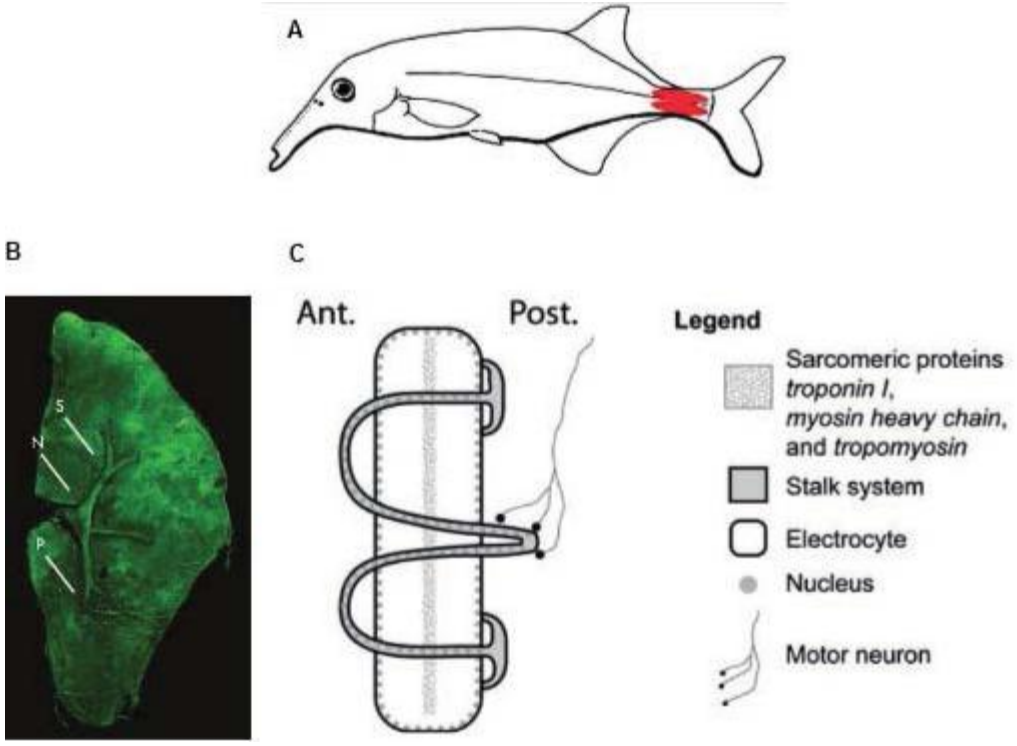


Figure 3 Electric organ anatomy. **A)** Position of the electric organ (red ovals) in the caudal peduncle of a mormyrid fish (Hopkins 2009). **B)** Confocal reconstruction of a mormyrid's electrocyte, showing the stalk system (S) with the relative penetrations (P) and the innervation pattern (N) (Gallant *et al.* 2014). **C)** Schematic view of the same electrocyte (Gallant *et al.* 2012).

3.3 History of an adaptive radiation.

Adaptive radiation consists in the relatively rapid multiplication of evolutionary lineages from an ancestral taxon; this phenomenon is often accompanied by remarkable levels of phenotypic and ecological differentiation and it constitutes one of the most extensively investigated phenomena in evolutionary biology (Schluter 2000). Exemplary cases of adaptive radiations include: the Galapagos' finches Geospizinae (Petren *et al.* 2005), the neotropical *Anolis* lizards (Irschick *et al.* 1997) and the East African cichlids (Losos 2010). Adaptive radiation is the outcome of divergent natural selection coupled with the occurrence of isolation mechanisms that interrupt gene-flow, either totally or partially, among diversifying populations. Such isolation mechanisms may be the natural consequence of geographic separation (e.g. the colonization of different islands of an archipelago) and get strengthened, at a later stage, during secondary contact (i.e. reinforcement, Hoskin *et al.* 2005). In the case of sympatric populations, however, strong diversifying selection, for example associated to the

exploitation of different ecological niches, is typically related to increased levels of assortative mating and hybridization avoidance (Maynard Smith 1966).

African weakly electric fishes have undergone an outstanding adaptive radiation possibly owing to their ability to produce and perceive weak electric signals. As already stated, mormyrids alone include approximately the 90% of all extant osteoglossomorphs. The acquisition of electric organs and electroreceptors could have opened new ecological opportunities for these fishes acting, thus, as a key innovation. At the same time, the high levels of diversification observed in the EOD of mormyrids might have facilitated strong assortative mating and, therefore, promoted speciation. This hypothesis is further supported by the fact that the sister family of Mormyridae –Gymnarchidae– counts only one species: *Gymnarchus niloticus*. This species, although capable to produce a continuous wave-type EOD, is characterized by simple, "stalkless" electrocytes. Mormyrid fishes, on the other hand, possess highly variable electrocyte anatomies that are associated to different pulse-type EOD shapes and lengths. While both EOD types serve basically the same ancestral function, i.e., electrolocation (von der Emde 1999), pulse-type EODs can be additionally used in social communication (Gebhardt *et al.* 2012), species and mate recognition (Bratton & Kramer 1989; Feulner *et al.* 2009b). Therefore, the development of an "evolutionarily plastic" electrocyte may stand at the basis of the adaptive radiation of African weakly-electric fishes.

There are mainly two ways by which EODs could promote speciation in mormyrid fishes: they can act as pre-zygotic isolation mechanisms following adaptation to different environments (ecological speciation, Nosil 2012) or they can directly trigger speciation by sexual selection (Panhuis *et al.* 2001). The latter mechanism seems to be important for species displaying marked sexual dimorphism in terms of EOD shape and duration, like representatives of the genus *Paramormyrops* (Arnegard *et al.* 2010). All the so-far investigated *Campylomormyrus* species, instead, seem to show no difference between the two sexes' EODs (Feulner *et al.* 2009a), hence suggesting ecological speciation as the main modality of diversification (Feulner *et al.* 2008). It is worth mentioning here that another taxon of weakly-electric fishes: the phylogenetically unrelated South American Gymnotiformes, have undergone a similar adaptive radiation (see Lavoué *et al.* 2012). This striking case of convergence seems to sustain the role of key innovation played by the EOD during adaptive radiations.

3.4 Aims of this study.

Analysis of transcriptomic differences between skeletal muscle and electric organ.

Despite the importance of the EOD as a potential trigger for adaptive radiations, very few studies have so far tried to investigate the genetic basis of its production and evolution in African weakly-electric fish (Gallant *et al.* 2012). Addressing this kind of question requires the comprehension of the role played by regulatory networks in determining the phenotypic differences observed between

skeletal muscle and electric organ and, at a later stage, across phenotypically divergent electric organs. This task, however, is particularly challenging for non-model organisms –i.e., species lacking extensive genomic resources– like mormyrid fishes.

The development of inexpensive 'new-generation' sequencing technologies has released an enormous potential in terms of sequence data generation, leading to the use of large scale genomic tools in molecular ecology and evolutionary studies. In particular, the sequencing of transcriptomes –i.e., the transcribed portion of the genome– results to be a cost-effective strategy, especially for non-model organisms (Ekblom & Galindo 2011).

The major aim of this study is to identify, describe and quantitatively assess the transcriptome-wide patterns of gene expression between skeletal muscle and electric organ within mormyrid fishes. The approach adopted for tackling this problem is transcriptome sequencing via mRNA-Seq, a technology that leverages the power of Next-Generation Sequencing (NGS) to reveal a snapshot of messenger RNA –i.e., the fraction of total RNA coding for proteins– presence and quantity in a certain tissue at a given moment in time (Wang *et al.* 2009).

Campylomormyrus was chosen as a model taxon for mainly two reasons: i) unlike other mormyrids, its representatives do not show any evident sign of sexual dimorphism or seasonal variability, facilitating thus inter-specific comparisons; and ii) because it is relatively easy to rear, breed and, in some cases, cross-breed in captivity. The combination of these two features make of *Campylomormyrus* a prospective model organism for future phenotype/genotype association studies regarding the EOD.

Molecular phylogeny of the genus Campylomormyrus.

For a proper understanding of the mutual relationships between speciation and ecological or phenotypic diversification, a robust phylogeny of the taxon under study is an essential prerequisite. A robust, well-supported phylogeny can in fact be employed for testing hypotheses regarding character evolution, ancestral state reconstruction and divergence dating. Genus-level phylogenies, however, are often unresolved at terminal branches due to lack of resolution in the employed markers. Until recently, it has been common practice to extrapolate species trees from gene trees obtained by utilizing one or a few concatenated loci, sampled from single individuals per species. This method, however, does not take into account possible discordances among gene trees, which may arise due to phenomena such as incomplete lineage sorting (ILS), occurrence of gene flow among diverging units, hybridization, and gene duplication/extinction, that may hence lead to the estimation of incorrect species trees (Pamilo & Nei 1988; Kubatko & Degnan 2007).

In the present study I aim at producing a reliable species-tree for the genus *Campylomormyrus* using methods based on the multispecies coalescent model; a model that bridges together classic phylogenetic inference with population genetics, by considering intra- and interspecific diversity and

by explicitly modelling the stochastic processes that generate discordance between species- and gene-trees (Heled & Drummond 2010).

A second task is to identify putatively cryptic species by applying coalescent-based models of species delimitation (Fujita *et al.* 2012) and by using microsatellite markers for the detection of subtle differences among supposed species.

4 Summary of articles

4.1 Summary of article I:

Lamanna, F., Kirschbaum, F. & Tiedemann, R. 2014

De novo assembly and characterization of the skeletal muscle and electric organ transcriptomes of the African weakly electric fish *Campylomormyrus compressirostris* (Mormyridae, Teleostei).

Molecular Ecology Resources **14**, 1222-1230.

Despite the importance of the electric organ discharge (EOD) as an effective prezygotic isolation mechanism, the investigation of its genetic bases remained so far limited. In this study I produced a first draft of the skeletal muscle and electric organ transcriptomes from the weakly electric fish species *Campylomormyrus compressirostris*, obtained using the Illumina HiSeq2000 sequencing technology. At the time of publication, they were the first transcriptomes available for a weakly-electric fish. Approximately 6.8 Gbp of cDNA sequence data were produced from both tissues, resulting in 57 268 109 raw reads for the skeletal muscle and 46 934 923 for the electric organ, and assembled *de novo* into 46 143 and 89 270 contigs, respectively. About 50% of both transcriptomes were annotated after protein databases search. The two transcriptomes show similar profiles in terms of Gene Ontology (GO) functional categories composition. Several candidate genes, which are likely to play a central role in the production and evolution of the electric signal, were identified and annotated. For most of these genes, and for many other "housekeeping" genes, I was able to obtain the complete or partial coding DNA sequences (CDS), which can be used for the development of primers to be utilized in future qRT-PCR experiments. I present also the complete mitochondrial genome and compare it to those available from other weakly electric fish species. Additionally, I located 1671 Short Sequence Repeat (SSR)-containing regions as well as their flanking sites and designed the relative primers. This study establishes a first step towards the development of genomic tools aimed at understanding the role of electric communication during speciation.

Authors' contributions:

I performed all the lab work, the bioinformatic analyses and wrote the manuscript. R. Tiedemann supervised the study and together with F. Kirschbaum took part in the discussion. All authors read and provided advice on the manuscript.

4.2 Summary of article II

Lamanna, F., Kirschbaum, F., Waurick, I., Dieterich, C. & Tiedemann, R.
Transcriptome-wide analysis of gene expression between skeletal muscle and electric organ in two species of African weakly-electric fish (Teleostei; Mormyridae; *Campylomormyrus*).
BMC Genomics, submitted.

In this study, I aimed at exploring the constitutive differences in terms of gene expression across two tissues: the electric organ (EO) and skeletal muscle (SM) in two species of the African weakly-electric fish *Campylomormyrus* (*C. compressirostris*, *C. tshokwe*), by using mRNA-seq based differential expression (DE) analysis.

The tissue samples utilized in this study were collected during my fieldwork in the Republic of the Congo in August/September 2012.

Eight paired-end (100bp), strand-specific, Illumina libraries were sequenced, producing 237,857,650 quality-filtered short read pairs. The obtained reads were assembled *de novo* into two reference transcriptomes (one for each species). *In silico* DE-analysis allowed us to identify 267 shared differentially expressed genes between EO and SM in the two species. For *in silico* DE-analysis is intended: i) the alignment of the Illumina reads onto the reference transcriptome; ii) the abundance quantification of the aligned reads using normalized counts, namely Fragments per kilobase per million reads mapped (FPKM); iii) the identification of differentially expressed genes between the two tissues by modelling intra-condition and inter-condition variance.

The obtained results suggest that: i) the loss of contractile activity and the decoupling of the excitation-contraction processes are reflected by the down-regulation of the corresponding genes in the electric organ; ii) the metabolic activity of the EO might be specialized towards the production and turn-over of membrane structures; iii) several ion channels are highly expressed in the EO in order to increase excitability; iv) several myogenic factors might be down-regulated by transcription repressors in the EO.

Authors' contributions:

I participated in specimen sampling, extracted the RNA, carried out the bioinformatic analyses and wrote the manuscript. F. Kirschbaum participated in specimen sampling, determined sex and species for the collected samples and participated in manuscript drafting and supervision. I. Waurick produced the cDNA libraries and prepared them for sequencing. C. Dieterich participated in experimental design and provided sequencing resources. R. Tiedemann participated in specimen sampling and supervised the study. All authors read and provided advice on the manuscript.

4.3 Summary of article III

Lamanna, F., Kirschbaum, F., Schneider, A. R. R., Feulner, P. G. D., Paul, C., Mamonekene, V. & Tiedemann R. 2014.

Coalescent-based Species Tree Estimation and Delimitation in a Group of Sympatric Weakly-Electric Fish (Osteoglossiphormes, Mormyridae, *Campylomormyrus*).

The principal aim of this study was to resolve the phylogenetic relationships between members of the genus *Campylomormyrus* by applying methods based on the multispecies coalescent model (Heled & Drummond 2010). Such model bridges together classic phylogenetic inference with population genetics, by considering intra- and interspecific diversity and by explicitly modelling the stochastic processes that generate discordance between species- and gene-trees. Five molecular markers (one mitochondrial and four nuclear genes) were used on a dataset composed of 15 taxonomic units (10 already described species and 5 putative new species based on morphological and electric signal characteristics). The goodness-of-fit of the obtained species-tree to the multispecies coalescent model was further assessed by using Posterior Predictive Simulations (PPS): a statistical technique which allows to compare an empirical species-tree with a set of trees simulated under the multispecies coalescent model.

Several coalescent-based species delimitation models were then applied on the obtained species-tree and their results were compared. Species delimitation was additionally investigated using 16 microsatellite loci as an independent line of evidence.

The results obtained in this study allowed me to disentangle the evolutionary relationship between two clearly distinct morphological species (*C. compressirostris* and *C. curvirostris*), which were not resolved by previous phylogenetic analyses (Feulner *et al.* 2007). Additionally, the analyses confirm the genetic isolation of two individuals (C01, K15) that are morphologically different from all available *Campylomormyrus* holotypes, thus confirming the presence of a new, yet undescribed species.

Authors' contributions:

I performed specimen sampling, most of the lab work, analyzed the data and wrote the manuscript. F. Kirschbaum participated in specimen sampling and species identification. A.R.R. Schneider helped with the lab work. P.G.D. Feulner provided several DNA sequences and microsatellite genotypes. C. Paul participated in specimen sampling and primer design. V. Mamonekene participated in specimen sampling and provided logistic support. R. Tiedemann participated in specimen sampling and took part in the discussion together with F. Kirschbaum. All authors read and provided advice on the manuscript.

5 Article I

De novo **assembly and characterization of the skeletal muscle and electric organ transcriptomes of the African weakly electric fish** ***Campylomormyrus compressirostris*** **(Mormyridae, Teleostei).**

De novo assembly and characterization of the skeletal muscle and electric organ transcriptomes of the African weakly electric fish Campylomormyrus compressirostris (Mormyridae, Teleostei)

FRANCESCO LAMANNA ,* FRANK KIRSCHBA UM † and R A L P H T I E D E M A N N *

*Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, Potsdam, Germany, †Department of Crop and Animal Sciences, Faculty of Horticulture and Agriculture, Humboldt University of Berlin, Philippstrasse 13, Berlin, Germany

Abstract

African weakly electric fishes (Mormyridae) underwent an outstanding adaptive radiation (about 200 species), puta-tively owing to their ability to communicate through species-specific weak electric signals. The electric organ discharge (EOD) is produced by muscle-derived electrocytes organized in piles to form an electric organ. Despite the importance of this trait as a prezygotic isolation mechanism, genomic resources remained limited. We present here a first draft of the skeletal muscle and electric organ transcriptomes from the weakly electric fish species Camp-ylomormyrus compressirostris, obtained using the Illumina HiSeq2000 sequencing technology. Approximately 6.8 Gbp of cDNA sequence data were produced from both tissues, resulting in 57 268 109 raw reads for the skeletal muscle and 46 934 923 for the electric organ, and assembled de novo into 46 143 and 89 270 contigs, respectively. About 50% of both transcriptomes were annotated after protein databases search. The two transcriptomes show similar profiles in terms of Gene Ontology categories composition. We identified several candidate genes which are likely to play a central role in the production and evolution of the electric signal. For most of these genes, and for many other housekeeping genes, we were able to obtain the complete or partial coding DNA sequences (CDS), which can be used for the development of primers to be utilized in qRT-PCR experiments. We present also the complete mitochon-drial genome and compare it to those available from other weakly electric fish species. Additionally, we located 1671 SSR-containing regions with their flanking sites and designed the relative primers. This study establishes a first step in the development of genomic tools aimed at understanding the role of electric communication during speciation.

Keywords: cSSR, electric fish, electric organ, skeletal muscle, transcriptome

Received 23 January 2014; revision received 18 March 2014; accepted 24 March 2014

Introducion

Understanding the genetic basis of speciation constitutes a pivotal aim in modern evolutionary biology. During the last 20 years, the investigation of ‘speciation genes’, that is, genes involved in the onset of

reproductive isolation, has attracted increasing attention within the scientific community (Noor & Feder 2006; Nosil & Schluter 2011; Butlin et al. 2012). Research on speciation genes addresses the question about the role played by individual genes and gene networks during the speciation process. However, before next-generation sequencing technologies became available, such efforts were restricted to few model organisms, excluding from research many groups of species which were historically recognized as interesting 'speciation models', but for whom little information regarding their genetics was available. One of these groups is the Mormyridae, an endemic family of African weakly electric freshwater fish. With at least 188 described species (Daget et al. 1991; Alves-Gomes & Hopkins 1997), this family accounts for about 90% of species diversity within the basal teleost taxon Osteoglossomorpha (bony tongue fish). Mormyrids are differentiated from other Osteo- glossomorphs by the capability to produce a weak, pulse-like, electric organ discharge (EOD), which allows them to localize objects in murky waters, by measuring alterations in their self-produced electric field (active electrolocation; Bullock et al. 2005). Additionally, EOD plays an essential role in social communication, species recognition and reproduction (Arnegard et al. 2010; Carl-son et al. 2011). In adult individuals, the EOD is pro-duced by a specific electric organ, consisting of modified muscle cells (i.e. electrocytes) and is perceived by two classes of tuberous electroreceptors (i.e. mormyromasts and knollenorgans; Bullock et al. 2005).

Mormyrids can discriminate between the EODs of conspecifics and heterospecifics and between those of females and males, relying on temporal and shape pat-terns of the EOD waveform (Feulner et al. 2009a). Modifi- cations of these two features–shape and duration–appear to act as effective prezygotic isolation mechanisms and are likely to have promoted the radiation of mormyrid fish to such a high number of species (Feulner et al. 2008,2009b).

Since the beginning of the second half of the twentieth century, the histology, anatomy and neurophysiology of EOD production have been thoroughly analysed (Liss- mann 1951, 1958). These and subsequent studies revealed a complex phenotypic architecture behind the production and modification of EODs. In fact, there are at least four factors which could influence EOD parame- ters: (i) innervation patterns of single electrocytes (Bullock et al. 2005); (ii) distribution of ion channels on the cell surface (Gallant et al. 2012); (iii) modifications on active sites of ion channels; and (iv) differences in their expression levels across different tissues (Zakon et al. 2006). Despite the large amount of information gathered in the last years, no large-scale molecular data sets were produced yet. The development of inexpensive 'next-generation' sequencing technologies released an enormous poten- tial in terms of sequence data generation, leading to the use of large scale genomic tools in molecular ecology and evolutionary studies. In particular, the sequencing of transcriptomes results to be a cost- effective analysis strategy, especially for nonmodel organisms (Vera et al. 2008; Elmer et al. 2010; Ekblom & Galindo 2011). We present here a first version of the transcriptomes of the electric organ and the skeletal muscle from the African weakly electric fish species Campylomormyrus compressirostris, using the Illumina HiSeq2000 sequencing technology. This particular species was chosen for its rel-ative ease of rearing and breeding in captivity, compared to other mormyrids, allowing researchers to obtain in future an extensive amount of information about its physiology, development, behaviour and

genetics. Illu- mina HiSeq2000 is currently one of the most widely used platforms for transcriptome sequencing, due to its high throughput (~100 millions of reads per lane) and the recent increase in read length (up to 150 bp; Crawford et al. 2010; Wang et al. 2010; Feldmeyer et al. 2011; Garg et al. 2011; Iorizzo et al. 2011; Fox et al. 2013; Wu et al. 2013; Zhang et al. 2013). The aims of the present study are to (i) characterize the skeletal muscle and electric organ transcriptomes of an African weakly electric fish species, by checking for differences in functional category composition across the two tissues; (ii) identify a pool of candidate genes which might be involved in the determination of the species specificity of the EOD; and (iii) obtain sequence resources for the development of markers to be used in future population genetics and evolutionary studies (cSSRs, mtDNA).

Materials and methods

RNA isolation and cDNA synthesis

Muscle and electric organ tissues were dissected from a male specimen of Campylomormyrus compressirostris after anesthetization with clove oil and euthanasia. The dis-sected tissues were first frozen in liquid nitrogen, ground with precooled mortar and pestle and additionally homogenized with a Polytron PT 1200 E homogenator.

Total RNA extraction was performed using the RNeasy® Mini Kit (Qiagen), RNA quality and concen-tration were estimated using a NanoDrop 1000 spectrophotometer, and cDNA was produced through the MINT-Universal cDNA synthesis kit (Evrogen). Firststrand cDNA was generated with the CDS-3M adapter 50 – AAGCAGTGGTATCAACGCAGAGTGGCCGAGGC GGCC(T)20VN-30 and PlugOligo-3M adapter 50 – AAG CAGTGGTATCAACGCAGAGTGGCCATTACGGCCG GGGG-P-30 ; the full cDNA synthesis was performed with the PCR Primer M1 50 -AAGCAGTGGTATCAA CGCAGAGT-30 . A clean-up of the cDNA for subse-quent Illumina sequencing was performed with the NucleoSpin® Extract II Kit (Macherey-Nagel).

The research followed internationally recognized guidelines and applicable national legislation. We received ethical approval from the deputy of animal wel-fare of the University of Potsdam.

Transcriptome sequencing and assembly

The two separate non-normalized cDNA libraries were sent to an external company (LGC Genomics GmbH, Ber-lin, Germany) for transcriptome sequencing. The sequencing library was generated using the TruSeq RNA Sample Prep Kit (Illumina Inc.). Sequencing was per-formed using 1 channel of a HiSeq2000 Illumina Sequencing System (Single-end, 100 bp).

After the sequencing process, the resulting fastq files were quality-checked using the FastQC v0.9.2 software; reads were processed in a 4-step way: (i) removal of adapter sequences; (ii) removal of all reads

containing the ‘N’ character; (iii) trimming of reads at 30 -end to get a minimum average Phred quality score of 20 over a window of ten bases; and (iv) removal of reads shorter than 35 bases after trimming. This processing steps increase the quality of reads because they remove those reads/bases with sequencing faults, which occur on Illumina Systems mainly at the 30-end.

Two different software packages were used for contig assembly: Trinity (release 2011-10-29; Grabherr et al. 2011) and the Velvet (v1.2.01)/Oases (v0.2.06) pipeline (Zerbino & Birney 2008; Schulz et al. 2012). Both assem-blers are designed for short-read sequencing technolo-gies and rely on De Bruijn graphs to build the contigs (Pevzner et al. 2001). In this kind of graph, each node is defined by a sequence of fixed nucleotide numbers (k-mer), and a connection between nodes is determined by overlaps of k-1 nucleotides. With Velvet/Oases, it is possible to use several different k-mer lengths, while with Trinity, it is only possible to use one fixed value. The following k-mer values were used: Trinity, 25 (default value); Velvet/Oases, 33, 41, 51, 53, 57. The five contig sets resulting from the five different k-mer lengths were successively merged into a single assembly, using the command –merge = yes in Oases. K-mer lengths were chosen in order to maximize the N50 value and the amount of assembled reads in the final merged assembly.

Transcriptome annotation

There are so far no genomic data available for any spe-cies of the Osteoglossomorpha; therefore, we compared both assembled transcript sets to the zebrafish (Danio re-rio), nile tilapia (Oreochromis niloticus), three-spined stick-leback (Gasterosteus aculeatus), medaka (Oryzias latipes) and green-spotted pufferfish (Tetraodon nigroviridis) mRNA databases available from NCBI (January 2012 version). Similarity searches were conducted using the stand-alone versions of blastn and tblastx (Altschul et al. 1990), implemented in Blast+ (v2.2.25; Camacho et al. 2009), with an E-value cut-off of 10- 6. To identify the amount of unique transcripts covered by the two tran-scriptomes, an additional blastn comparison was per-formed against the Unigene database of D. rerio (build #126, 19 Jun 2012). The two sets of transcripts were also compared against the Swiss-Prot database (release-2012_01), using the blastx tool with an E-value cut-off of 10-10. Swiss-Prot is a high quality, manually curated nonredundant protein database, unlike electronically annotated databases (like TrEMBL). Swiss-Prot provides for each entry also information about the function of the corresponding protein, a very helpful feature when look-ing for candidate genes.

Sequences with positive matches on Swiss-Prot were further searched for Gene Ontology (GO) terms (Ashburner et al. 2000) with the software Blast2GO (Conesa et al. 2005), command-line version (B2G4pipe v2.4.0). Gene ontology terms constitute a controlled vocabulary of gene attributes. Each GO term is organized hierarchi-cally into three main distinct categories (‘molecular func-tion’, ‘biological process’, ‘cellular component’). Terms within each top-level category are connected by parent– child relationships, building a directed acyclic graph (DAG), in which each child term can be connected to one or more parent terms. GO IDs can be categorized on the basis of a smaller set of high-level terms (GO-slim). We used the GO-slim vocabulary v1.2 and the web tool cateGOrizer (http://www.animalgenome.org/tools/catego/) to

group the low level GO terms found in our data sets. The grouped sets of GO-slim were then subjected to a Fisher's exact test, using a custom script in (R Foundation for Statistical Computing 2012), using FDR P-value correction for multiple comparisons, in order to find under-/over-represented terms between the two transcriptomes. The FDR test controls the false discovery rate, that is, the expected proportion of false discoveries among the rejected hypotheses.

Candidate genes analysis

After Gene Ontology annotation, both transcript sets were searched for the following GO terms: 'Muscle System Process', (GO:0003012); 'Ion Channel Activity', (GO:0005216); 'Embryonic Development', (GO:0009792); and 'Axon Guidance', (GO: 0007411), which may contain candidate genes relevant for the development and/or function of a tissue as either a functional muscle or an electric organ.
All corresponding transcripts were extracted from our database and re-aligned to the mRNA cross-reference corresponding to their relative top blast-hit in SwissProt using MAFFT v7.017 (Katoh et al. 2002). We annotated the aligned portion of each of our transcripts with a puta-tive coding DNA sequence (CDS) and discarded all the sequences with internal stop codons. RPKM values were calculated for each candidate gene in the two transcripto-mes using RSEM (v1.2.3; Li & Dewey 2011).

mtGenome reconstruction

We reconstructed the C. compressirostris complete mito-chondrial genome by mapping our reads back to the almost complete mitochondrial genome sequence of the closely related species Campylomormyrus numenius (Lavou'e et al. 2012) using the 'Map to reference' function implemented in Geneious v6.1.5 (http://www.geneious. com/).

SSR detection

We identified all SSRs (Small Sequence Repeats) in our data set, with a repeat length ranging from 2 to 6 and a minimum sequence length of 12nt, using the program Phobos v3.3.12 (Mayer 2006–2010). Primer couples were designed for all the retrieved SSRs using the program BatchPrimer3 v1.0 (You et al. 2008).

Results

Transcriptome sequencing and assembly

Sequencing of the two cDNA libraries on the Illumina HiSeq2000 platform produced 34 963 283 and 43 097 741 quality-filtered reads for electric organ and skeletal muscle, respectively (Table 1). Assembly results

obtained from Trinity and Velvet/Oases are compared in Table 2 and Fig. 1. Two parameters were taken into consideration when assessing the quality of the two assemblies: N50 and the proportion of annotated contigs after a blast search against the SwissProt (Table 2) and Unigene (Fig. 1) databases. In both cases, Velvet/Oases gave better results than Trinity, and the resulting assembly was chosen for all downstream analyses. The Velvet/Oases assembly produced 89 270 contigs for the electric organ and 46 143 for the skeletal muscle.

Annotation

About 10.6% and 11.1% sequences, from the electric organ and muscle transcriptomes, respectively, matched against the Danio rerio mRNA record after a blastn search. With the tblastx tool, we obtained about 46.6% and 46.9% of matches against the D. rerio database. This dramatic increase in the percentage of matches is explained by the fact that protein sequences are much more conserved than nucleotide sequences; this effect is particularly pronounced when comparing distantly related species, which belong to different superorders, like in the case of Campylomormyrus compressirostris (Os-teoglossmorpha) and D. rerio (Elopocephala). A similar pattern emerged when our data were compared to the

Table 1 Sequencing statistics for Illumina run (one lane, single-end, 100 bp)

	EO (Velvet/ Oases)	EO (Trinity)	MU (Velvet/ Oases)	MU (Trinity)
No. of contigs	89 270	60 654	46 143	33 562
No. of assembled bp (n)	83 738 964	39 520 374	42 047 650	20 207 955
N50 (bp)	1422	992	1319	847
GC content (%)	46.24	46.31	46.1	46.05
Matches in SP (%)	47.5	42.9	49.6	44.6

*Standard deviation in parentheses.

Table 2 Comparison of the annotation results obtained from Trinity and Velvet/Oases

	Electric organ (EO)	Muscle (MU)
Reads raw (n)	46 934 923	57 268 109
Reads processed (n)	34 963 283	43 097 741
Bases (bp)	3 063 383 265	3 792 824 916
Average read length (bp)*	86.38 (19.91)	86.76 (19.92)

Figure 1

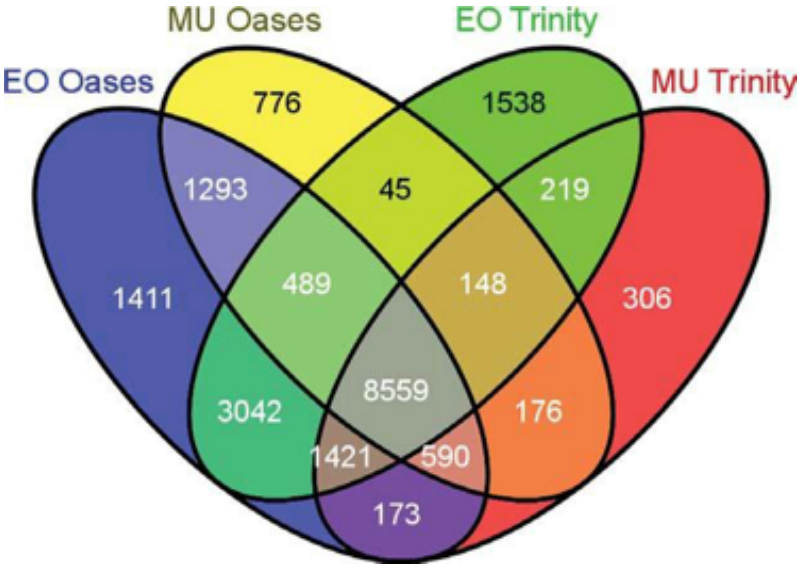


Fig. 1 Venn diagram showing the overlap in terms of numbers of blastn matches against the Danio rerio Unigene database, for the electric organ (EO) and skeletal muscle (MU) transcriptomes assembled with Trinity and Velvet/Oases. Amount of matched sequences for each assembly: EO Oases = 16 978; MU Oases = 12 076; EO Trinity = 15 461; MU Trinity = 11 592.

reference databases of G.aculeatus, Oreochromis niloticus, Oryzias latipes, Tetraodon nigroviridis (Fig. 2). The comparison against the D. rerio Unigene database yielded 16 978 and 12 076 positive matches for the electric organ and skeletal muscle transcriptomes, respectively. In total, 42 422 (47.5%) and 22 895 (49.6%) sequences from the electric organ and muscle transcriptomes showed positive matches when compared to the SwissProt database. They were further annotated using the GO vocabulary: 40 143 (94.6%) sequences from the electric organ data set and 21 826 (95.3%) from the skeletal muscle could be annotated with GO terms.
In order to look for under- or over-represented GO categories between the two transcriptomes, a Fisher's

Figure 2

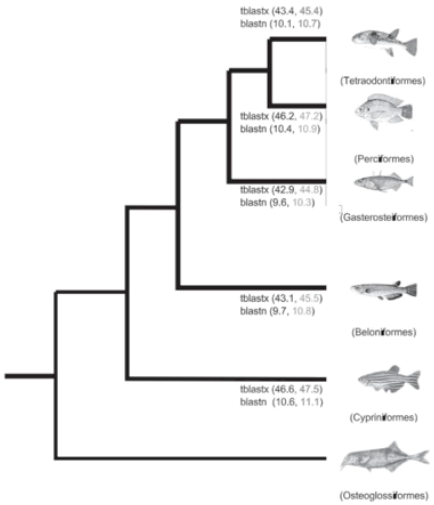


Fig. 2 Phylogenetic relationships of Campylomormyrus compressi-rostris with the compared fish species. Numbers in parentheses indicate the percentage of matching transcripts between the C. compressirostris electric organ (first value; black) and skeletal muscle (second value; grey) transcriptomes against the mRNA databases for every given species.

exact test with correction for multiple P-values was per-formed, using generic GO terms (GO slims) as a refer-ence. The analysis revealed that the GO composition of the two transcriptomes is very similar, with none of the terms being over- or under-represented in either tissue (Fig. S1, Supporting information).

Candidate gene analysis

We identified 45 transcripts with complete or partial CDS corresponding to the GO terms 'muscle system pro-cess', 'ion channel activity', 'embryonic development' or 'axon guidance' in both transcriptomes. The sequences relative to the last three terms correspond to genes poten-tially involved in the production and modification of the EOD (Table S1, Supporting information), while those cor-responding to the first term are all genes devoted in maintaining the structural and functional architecture of the skeletal muscle (Table S2, Supporting information).

mtGenome reconstruction

An amount of 4 265 022 reads from the electric organ (EO) transcriptome mapped against the partial mitochondrial genome of Campylomormyrus numenius. The mapped reads spanned over a 100 bp long gap in the C. numenius control region. The completeness of the retrieved mt genome was additionally confirmed by aligning it to a set of complete mt genomes from related genera (see Supporting Information).
The mitochondrial genome of C. compressirostris shows a length of 16 715 bp (Fig. 3). It is the first complete mitochondrial genome for a representative of the genus Campylomormyrus and shows a high nucleotide similarity (>90%) to the closely related genera Gnathonemus, Genyomyrus and Cyphomyrus (Table 3).

SSRs detection

More than 35 000 SSRs sequences were detected in the two assembled transcriptomes. However, in order to get only unique sequences, primers were designed only from those transcripts which matched against the Uni-gene database; in this way, we obtained 1671 primer cou-ples (Table S3, Supporting information).

Discussion

Electric organs have fascinated naturalists for a longtime; even Darwin in the 'The Origin of Species' described them as striking examples of convergent evolution (Darwin 1859) and, since then, many studies were concerned with the evolution of these organs and the lineages of fishes presenting this feature.

The current study describes the first large-scale transcriptome sequencing of a weakly electric fish. It demonstrates the possibility to sequence, assemble and annotate a transcriptome de novo for a nonmodel organism like Campylomormyrus compressirostris, using only single-end, short-read sequencing technologies.

The number of generated raw reads (~113 mio.) is in line with the average output obtained using one lane of an Illumina Hiseq2000 platform. The amount of matches against the Unigene database, 16 978 (electric organ) and 12 076 (skeletal muscle), suggests a good coverage in terms of transcriptome diversity, despite the use of a non-normalized library. The different number of sampled genes is probably explained by the fact that in muscle tissue, there are several specific components – for example, myosin, creatine kinase – that are massively expressed in order to maintain the functional role of the organ; when using a non-normalized library, these highly expressed genes account for a huge number of reads, which can prevent the detection of rarer transcripts. The specieswise blastn and tblastx comparisons showed the highest number of matches when applied to the D. rerio set of transcripts (Fig. 2), probably because its genome is much more comprehensively

Figure 3

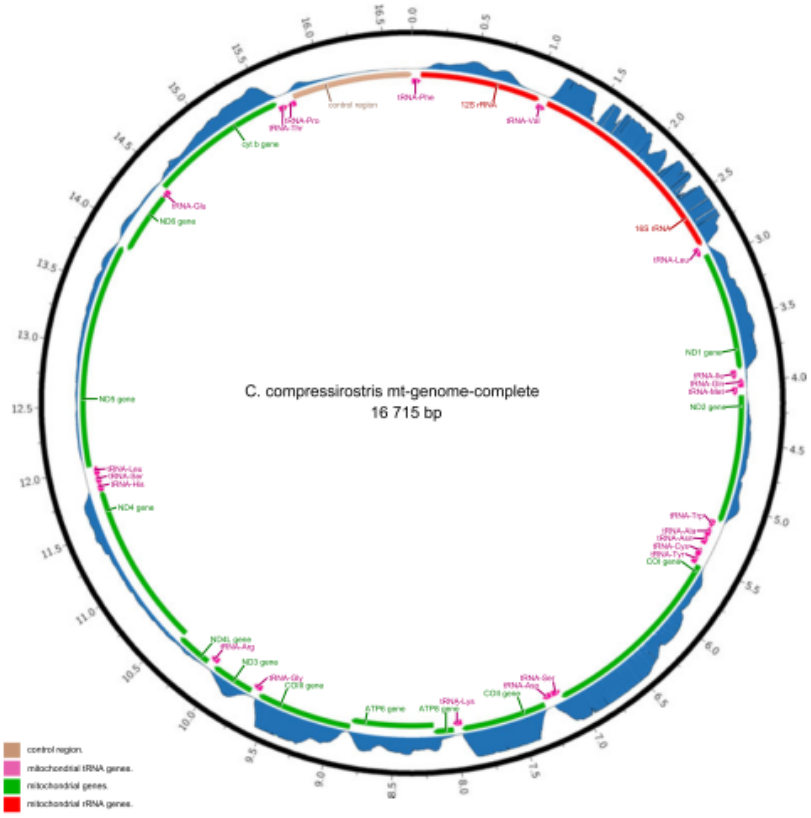


Fig. 3 Circular representation of the Campylomormyrus compressirostris mitochondrial genome with relative annotations. The blue histo-gram indicates coverage depth at each position after reads mapping (max coverage = 26 2959). Numbers on the outermost circle indi-cate length in thousands of bp.

annotated than those of the other compared species. The amount of assembled transcripts and the proportion of them that were annotated against the Swiss-Prot database (~50%) are consistent with the results obtained in other studies that applied Illumina de novo transcriptome sequencing to nonmodel fish species (Fraser et al. 2011; Fox et al. 2014; Schunter et al. 2014). The GO annotation reveals that the two transcriptomes have very similar profiles, in fact none of the gene ontology terms found in our data sets vary significantly between the two kinds of tissues (Fig. S1, Supporting information). However, it is worth pointing out that this kind of analysis was performed on a large scale and in a coarse fashion, in which thousands of transcripts were clustered into 116 high-level GO-terms. A more thorough inspection would probably identify more subtle differences. The presence of typical skeletal muscle transcripts (Table S2, Supporting information), found

Table3 Mitochondrial genome nucleotide similarity.

	Petrocephalus microphtalmus	*Myomyrus macrops*	*C. compressirostris*	*G. petersii*	*G. donnyi*	*C. discorhynchus*	*Brienomyrus niger*	*Paramormyrops gabonensis*
Petrocephalus microphtalmus	—							
Myomyrus macrops	81.7	—						
C. compressirostris	81.4	85.4	—					
Gnathonemus petersii	81.6	85.7	94.2	—				
Genyomyrus donnyi	81.8	86.0	92.9	93.2	—			
C. discorhynchus	81.8	85.7	92.8	92.9	93.9	—		
Brienomyrus niger	82.1	85.6	88.9	89.2	89.5	89.3	—	
Paramormyrops gabonensis	82.1	85.5	88.8	89.1	89.4	89.3	89.9	—

also within the electric organ transcriptome, indicates that these genes are still expressed in that tissue. This is not unexpected when considering that electrocytes derive from striated muscle. As electric organs constitute a noncontractile tissue, part of these transcripts are probably not translated into protein, invoking the role of post-transcriptional regulation mechanisms in the determination of the electric organ phenotype, as already demonstrated in studies on species belonging to the distantly related family Gymnotidae of weakly electric fish (Cuellar et al. 2006).

Candidate gene analysis revealed the presence of several genes which might play a role in the production and modification of the electric signal (Table S1, Supporting information). The identification of the CDS sequences for several of these candidate genes will help in developing primers for qPCR, which will be used to asses gene expression levels in other Campylomormyrus species, fea-turing different EODs and, hence, contributing to under-stand the role of gene expression during speciation.

Beside the discovery of functional units, transcripto-mes can be used for the development of genomic tools to be applied in population genetics analysis (e.g. SNPs, mi-crosatellites).

In this study, we obtained the whole mitochondrial genome sequence of C. compressirostris, and it includes also a small portion of the control region which should not be transcribed, because is located between the two transcription promoters of the mt genome (Shadel & Clayton 1997). The presence of this sequence in our transcriptomes might be due to a small amount of genomic DNA contamination in the sequenced libraries. The retrieved mt genome will be employed to better understand phylogenetic relationships in the genus Campylomormyrus and to analyse population differences within species.

Additionally, the identification of 1671 unique SSR containing regions, and the design of the relative primers, although not tested in the wet laboratory yet, might constitute a very valuable resource for future population genetics studies, which had been so far restricted to the only available 18 microsatellites (Feulner et al. 2005). As these sequences were obtained from cDNA data, the use of genomic DNA for testing the primers is discouraged, because many primer sequences might span exon boundaries and would likely not anneal to gDNA.

Acknowledgements

Funding was provided by the University of Potsdam and the Leibniz-SAW-project GENART. Computational resources were kindly provided by the Unit of Bioinformatics (J. Selbig, S. Hart-mann) at the University of Potsdam.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410.

- Alves-Gomes J, Hopkins C (1997) Molecular insights into the phylogeny of Mormyriform fishes and the evolution of their electric organs. Brain, Behavior and Evolution, 49, 324–351.
- Arnegard ME, McIntyre PB, Harmon LJ et al. (2010) Sexual signal evolu-tion outpaces ecological divergence during electric fish species radia-tion. The American Naturalist, 176, 335–356.
- Ashburner M, Ball CA, Blake JA et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet-ics, 25, 25–29.
- Bullock T, Fay R, Hopkins C, Popper A (2005) Electroreception. Springer Handbook of Auditory Research. Springer, New York, New York.
- Butlin R, Debelle A, Kerth C (2012) What do we need to know about spe-ciation?. Trends in Ecology & Evolution, 27, 27–39.
- Camacho C, Coulouris G, Avagyan V et al. (2009) BLAST+: architecture and applications. BMC Bioinformatics, 10, 421.
- Carlson BA, Hasan SM, Hollmann M et al. (2011) Brain evolution triggers increased diversification of electric fishes. Science, 332, 583– 586.
- Conesa A, Go€tz S, Garc'ıa-Go'mez JM, Terol J, Talo'n M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21, 3674–3676.
- Crawford JE, Guelbeogo WM, Sanou A et al. (2010) De novo transcrip-tome sequencing in Anopheles funestus using Illumina RNA-seq tech-nology. PLoS ONE, 5, e14202.
- Cuellar H, Kim JA, Unguez GA (2006) Evidence of post-transcriptional regulation in the maintenance of a partial muscle phenotype by elec-trogenic cells of S. macrurus. FASEB Journal, 20, 2540.
- Daget J, Gosse JP, Teugels GG, Thys van den Audenaerde DFE (1991) Check List of the Freshwater Fishes of Africa. ORSTOM, Paris, Bruxelles.
- Darwin C (1859) On the Origins of Species by Means of Natural Selection. Murray, London.
- Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107, 1–15.
- Elmer KR, Fan S, Gunter HM et al. (2010) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Molecular Ecology, 19, 197–211.
- Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M (2011) Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (Radix balthica, Basommatophora, Pulmo-nata), and a comparison of assembler performance. BMC Genomics, 12, 317.
- Feulner PGD, Kirschbaum F, Tiedemann R (2005) 18 microsatellite loci for endemic African weakly electric fish (Campylomormyrus, Mormyri-dae) and their cross species applicability among related taxa. Molecular Ecology Notes, 5, 446–448.

- Feulner P, Kirschbaum F, Tiedemann R (2008) Adaptive radiation in the Congo River: an ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; Campylomormyrus). Journal of Physiology, 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. Biology Letters, 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. Communicative & Integrative Biology, 2, 329–331.
- Fox SE, Preece J, Kimbrel JA et al. (2013) Sequencing and de novo tran-scriptome assembly of Brachypodium sylvaticum (Poaceae). Applications in Plant Sciences, 1, 1200011.
- Fox SE, Christie MR, Marine M et al. (2014) Sequencing and characteriza-tion of the anadromous steelhead (Oncorhynchus mykiss) transcrip-tome. Marine Genomics, (in press). Available from http://dx.doi.org/ 10.1016/j.margen.2013.12.001.
- Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA (2011) Sequencing and characterization of the guppy (Poecilia reticulata) transcriptome. BMC Genomics, 12, 202.
- 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. The Journal of Experi-mental Biology, 215, 2479–2494.
- Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identi-fication. DNA Research, 18, 53–63.
- Grabherr MG, Haas BJ, Yassour M et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Bio-technology, 29, 644–652.
- Iorizzo M, Senalik DA, Grzebelus D et al. (2011) De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. BMC Genomics, 12, 389.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research, 30, 3059–3066.
- Lavou'e S, Miya M, Arnegard ME et al. (2012) Comparable ages for the independent origins of electrogenesis in African and South American weakly electric fishes. PLoS ONE, 7, e36287.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformat-ics, 12, 323.
- Lissmann HW (1951) Continuous electrical signals from the tail of a fish, Gymnarchus niloticus Cuv. Nature, 167, 201–202.
- Lissmann HW (1958) On the function and evolution of electric organs in fish. Journal of Experimental Biology, 35, 156–191.

- Mayer C (2006–2010) Phobos 3.3.11. Available from http://www.rub.de/spezzoo/cm/cm_phobos.htm.
- Noor M, Feder J (2006) Speciation genetics: evolving approaches. Nature Reviews Genetics, 7, 851–861.
- Nosil P, Schluter D (2011) The genes underlying the process of speciation. Trends in Ecology & Evolution, 26, 160–167.
- Pevzner PA, Tang H, Waterman MS (2001) An Eulerian path approach to DNA fragment assembly. Proceedings of the National Academy of Sciences of the United States of America, 98, 9748–9753.
- R Foundation for Statistical Computing (2012) R: A Language and Environ-ment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from http://www. R-project.org.
- Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression lev-els. Bioinformatics, 28, 1086–1092.
- Schunter C, Vollmer SV, Macpherson E, Pascual M (2014) Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. BMC Genomics, 15, 167.
- Shadel GS, Clayton DA (1997) Mitochondrial DNA Maintenance in Verte-brates. Annual Review of Biochemistry, 66, 409–435.
- Vera JC, Wheat CW, Fescemyer HW et al. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology, 17, 1636–1647.
- Wang Z, Fang B, Chen J et al. (2010) De novo assembly and characteriza-tion of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (Ipomoea batatas). BMC Genomics, 11, 726.
- Wu CH, Tsai MH, Ho CC, Chen CY, Lee HS (2013) De novo transcrip-tome sequencing of axolotl blastema for identification of differentially expressed genes during limb regeneration. BMC Genomics, 14, 434.
- You FM, Huo N, Gu YQ et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinfor-matics, 9, 253.
- Zakon HH, Lu Y, Zwickl DJ, Hillis DM (2006) Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. Proceedings of the National Academy of Sciences of the United States of America, 103, 3675–3680.
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research, 18, 821–829.
- Zhang M, Li Y, Yao B et al. (2013) Transcriptome sequencing and de novo analysis for Oviductus Ranae of Rana chensinensis Using Illumina RNA-Seq Technology. Journal of Genetics and Genomics, 40, 137–140.

F.L. extracted the RNA, prepared the cDNA library for sequencing, assembled and annotated the data and drafted the manuscript. F.K. collected the sample and performed the dissection. R.T. conceived and supervised the study and participated in manuscript drafting. All authors read and approved the manuscript.

Data accessibility

The processed read sequences are available at the Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) under the Accession nos: SRR786741 (electric organ) and SRR797024 (skeletal muscle). The Accession nos for the candidate genes (KF860164– KF860208) are available at GenBank (http://www.ncbi.nlm.nih.gov/Genbank); the accession number for the mitochondrial genome (HG779437) is available at the European Nucleotide Archive (ENA) database (http:// www.ebi.ac.uk/ena). The contig files for the two transcriptomes, the results from the GO annotation, the mapping file of the reads on the Campylomormyrus numenius mt genome, the mt genomes alignment, and the R script used to perform the Fisher's exact test are available on Dryad (doi: 10.5061/dryad.ns135).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Gene Ontology (GO-Slim) terms composition of the elec-tric organ (EO) and Skeletal Muscle (MU) transcriptomes.

Table S1 List of genes potentially involved in electric organ dis-charge (EOD) generation and modification.

Table S2 List of housekeeping and skeletal muscle specific genes found in the two transcriptomes.

Table S3 Primer sequences, annealing temperatures, product size, repeat motif, number of repeats, expected product length, and sequence for 1671 cSSRs detected in the Campylomormyrus compressirostris transcriptomes.

6 Article II

Transcriptome-wide analysis of gene expression between skeletal muscle and electric organ in two species of African weakly-electric fish (Teleostei; Mormyridae; *Campylomormyrus*).

Transcriptome-wide analysis of gene expression between skeletal muscle and electric organ in two species of African weakly-electric fish (Teleostei; Mormyridae; *Campylomormyrus*).

Francesco Lamanna[1], Frank Kirschbaum[2], Isabelle Waurick[3], Christoph Dieterich[4,5], Ralph Tiedemann[1§].

[1]Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, Potsdam, Germany

[2] Department of Crop and Animal Sciences, Faculty of Horticulture and Agriculture, Humboldt University of Berlin, Philippstrasse 13, Berlin, Germany

[3]Museum für Naturkunde, Invalidenstraße 43, Berlin, Germany

[4]Max Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, Berlin-Buch, Germany

[5]Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Strasse 9B, Cologne, Germany

[§]Corresponding author

Email addresses:

FL: lamanna@uni-potsdam.de

FK: frank.kirschbaum@staff.hu-berlin.de IK: isabelle.waurick@mfn-berlin.de CD: christoph.dieterich@age.mpg.de RT: tiedeman@uni-potsdam.de

Abstract

Background

African weakly-electric fishes of the family Mormyridae are able to produce and perceive weak electric signals (≤ 1V) owing to the presence of a specialized, muscle-derived electric organ (EO) in their tail region. Such electric signals are used for objects/prey localization, for the identification of conspecifics, and in social and reproductive behaviour. The output of this organ, known as Electric Organ Discharge (EOD), might have promoted the adaptive radiation of this family by acting as an effective pre-zygotic isolation mechanism. Despite the physiological and evolutionary importance of this trait, the investigation of its genetic basis has so far remained limited. In this study, we aim at exploring constitutive differences in terms of gene expression across electric organ and skeletal muscle (SM) in two mormyrid species:
Campylomormyrus compressirostris and *C. tshokwe*, by using mRNA-Seq based differential expression (DE) analysis.

Results

Eight paired-end (100bp) strand-specific Illumina libraries were sequenced, producing 237,857,650 quality-filtered short read pairs. The obtained reads were assembled *de novo* in two reference transcriptomes (one for each species). *In silico* DE-analysis allowed us to identify 267 shared differentially expressed genes between EO and SM in the two species. We found that: i) several genes involved in cell membrane excitability (e.g., Na^+/K^+ pumps, voltage-gated Na^+ and K^+ channels) are up-regulated in EO; ii) several myogenic regulatory factors are down-regulated in the EO, whereas many of their known repressors and co-repressors (e.g., *hey1*, *hes6*) are up-regulated.

Conclusions

The data obtained indicate that: i) the loss of contractile activity and the decoupling of the excitation-contraction processes are reflected by the down-regulation of the corresponding genes in the electric organ's transcriptome; ii) the metabolic activity of the EO might be specialized towards the production and turn-over of membrane structures; iii) several ion

channels are highly expressed in the EO in order to increase excitability; iv) several myogenic factors might be down-regulated by transcription repressors in the EO.

Background

Bioelectrogenesis (i.e., the ability to produce strong or weak electric signals by specialized organs) has evolved several times independently in aquatic vertebrates [1]. It can be observed in the marine electric rays (Torpediniformes) and skates (Rajiformes), in the African freshwater Mormyridae and Gymnarchidae (Mormyroidea), in the South American knifefishes (Gymnotiformes), in several catfish species (Siluriformes), and in few marine stargazers (Perciformes). In all the above-mentioned groups, electric organs originate from myogenic tissue; the only exception are members of the family Apteronotidae (Gymnotiformes), where the electric organs are formed by modified spinal motor neurons [2]. The amount of excitable cells within each electric organ determines the electric potential of an EOD, which can range from few millivolts to several hundreds of volts (e.g., in the electric eel *Electrophorus electricus*).

African weakly-electric fishes of the family Mormyridae constitute a peculiar group of teleost fishes, formed by approximately 200 species [3], all endemic to African riverine and, partially, lacustrine systems. As their name suggests, they are able to produce only weak electric fields, in the order of millivolts to few volts, which are not used for predation or defence. The cells forming their electric organ are compressed disk-like cells commonly called electrocytes. In many species they are longitudinally stacked behind each other in order to form columns of cells embedded within tubes of isolating connective tissue. The synchronous activity of each electrocyte defines the output of the electric organ, known as Electric Organ Discharge (EOD) [4]. Such weak pulses are mainly employed for the localization and discrimination of objects in water (active electrolocation) [5], for the recognition of conspecific individuals [6, 7], and in social and reproductive behaviour [8, 9]. Mormyrids' EODs are considered to act as effective prezygotic isolation mechanisms, which have putatively facilitated ecological isolation among sympatric species and have, therefore, promoted the adaptive radiation observed in this family [10–13].

In all mormyrids, the adult electric organ is located in the caudal peduncle and is formed by four columns of electrocytes, two dorsal and two ventral ones. Each electrocyte is innervated by electromotoneurons originating in the spinal cord. Electric organs arise in juvenile fishes

from several myomeres of the deep lateral muscle; their myogenic origin is confirmed by the presence of disorganized myofibrils within the electrocytes [14, 15].

Despite the peculiarity of this organ and its importance for neurophysiology, animal communication and evolution, only very few studies have so far tried to investigate the genetic basis of its origins and functions [16–18]. The rise of the RNA-sequencing technology, coupled with the development of powerful algorithms and statistical methods, implemented in freely-available software packages, allows to reconstruct transcriptomes and to reliably assess their levels of expression, even for non-model organisms lacking extensive genomic resources, at a fraction of the cost and time necessary only few years ago [19, 20].

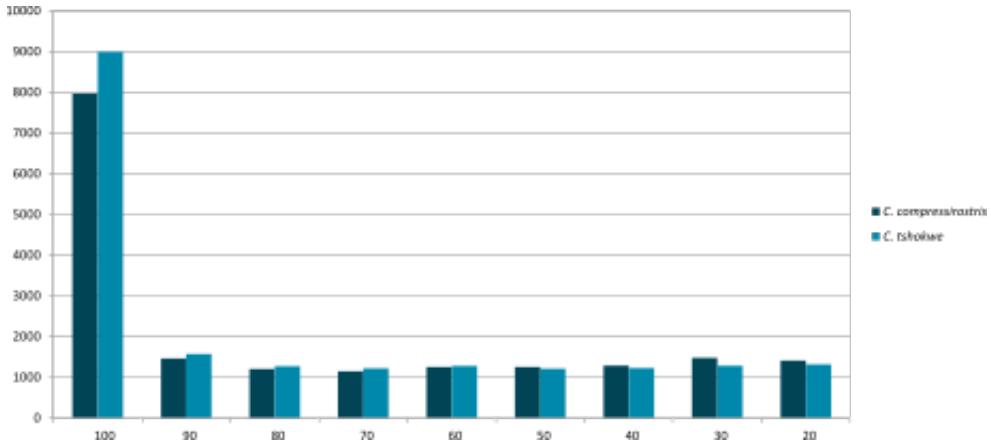
In the present work, we aim at exploring the differential patterns of gene expression between skeletal muscle (SM) and electric organ (EO) in adult specimens of *Campylomormyrus compressirostris* and *Campylomormyrus tshokwe*. We focus then on the identification of the differentially expressed genes that are in common between the two species and that might be responsible for the functional differences between the two tissues.These two species were chosen because they are relatively easy to rear, breed and cross-breed in captivity (F. Kirschbaum, personal communication), making them good candidates for future studies of ontogeny and genotype/phenotype association in weakly-electric fishes.

Results

Transcriptome sequencing and assembly

For each species, four paired-end (fragment length = 100bp), strand-specific cDNA libraries (two biological replicates per tissue) were sequenced on a single lane of an Illumina HiSeq2000 sequencing machine. Sequencing of the eight cDNA libraries produced a total amount of 267,052,872 raw read pairs, resulting in 237,857,650 quality-filtered read pairs (88%); see Additional file 1 for per library sequencing statistics. Trinity assembly resulted in 260,598 and 369,030 contigs for *C. compressirostris* and *C.tshokwe* transcriptomes respectively (Table 1). Contigs were then blasted against the *Danio rerio* proteome, retrieving 18,458 and 19,363 unique proteins for *C. compressirostris* and *C.tshokwe*, respectively; of these retrieved matches, 7971 (43.1%) and 8993 (46.4%) hits corresponded to full or nearly full-length coding sequences (Figure 1).

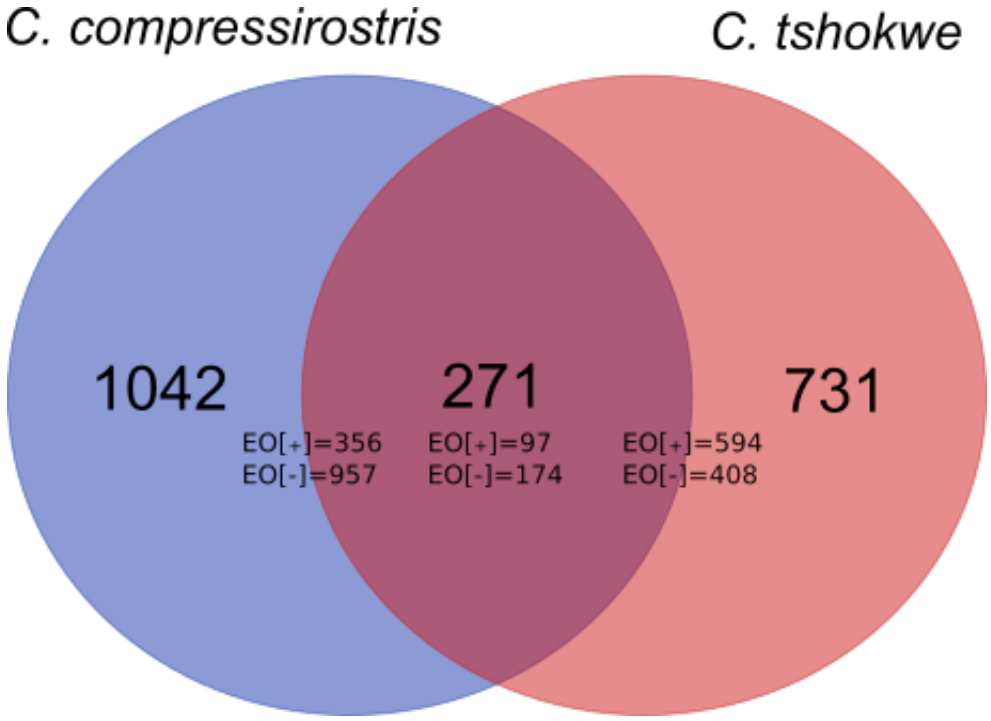
Figure 1



Differential Expression (DE) analysis

After transcript quantification with RSEM and DE-analysis with edgeR, 1313 transcripts resulted to be differentially expressed between EO and SM in *C. compressirostris* (356 up-regulated in EO and 957 down-regulated in EO) and 1002 in *C. tshokwe* (594 up-regulated in EO and 408 down-regulated in EO). Of all the differentially expressed transcripts, 267 resulted to be shared between the two species (97 up-regulated in EO and 171 down-regulated in EO) (Figure 2).

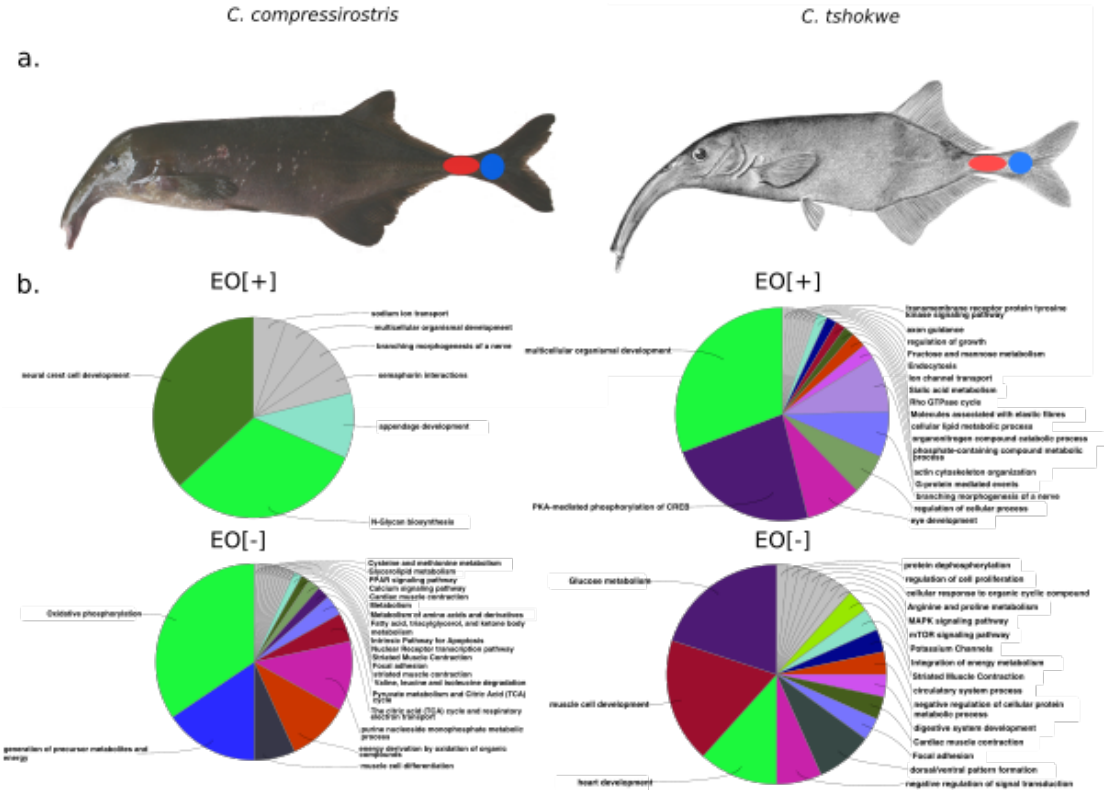
Figure 2



Functional annotation

In order to identify over-represented functional categories and pathways within the two transcriptomes, the sets of up- and down-regulated genes in the electric organ were subject to an enrichment analysis involving Gene Ontology categories and Reactome and KEGG biological pathways (Figure 3). The number of retrieved categories/pathways is proportional to the amount of differentially expressed genes present in each analysed set; with fewer categories for the EO up-regulated genes compared to the EO down-regulated genes in *C. compressirostris,* and a similar number of categories identified between the two sets in *C. tshokwe*. For both species the most represented categories within the EO up-regulated genes are related to ion channel transport (e.g., sodium ion transport), multicellular organismal development, and development/patterning of nerves (branching morphogenesis of a nerve; semaphorin interactions; axon guidance).The genes found to be down-regulated in the EO regard mainly functional classes like: metabolic pathways specific to muscle tissue (oxidative phosphorylation, pyruvate metabolism, calcium signalling pathway), and muscle specific categories (muscle cell differentiation, striated muscle contraction, cardiac muscle contraction).

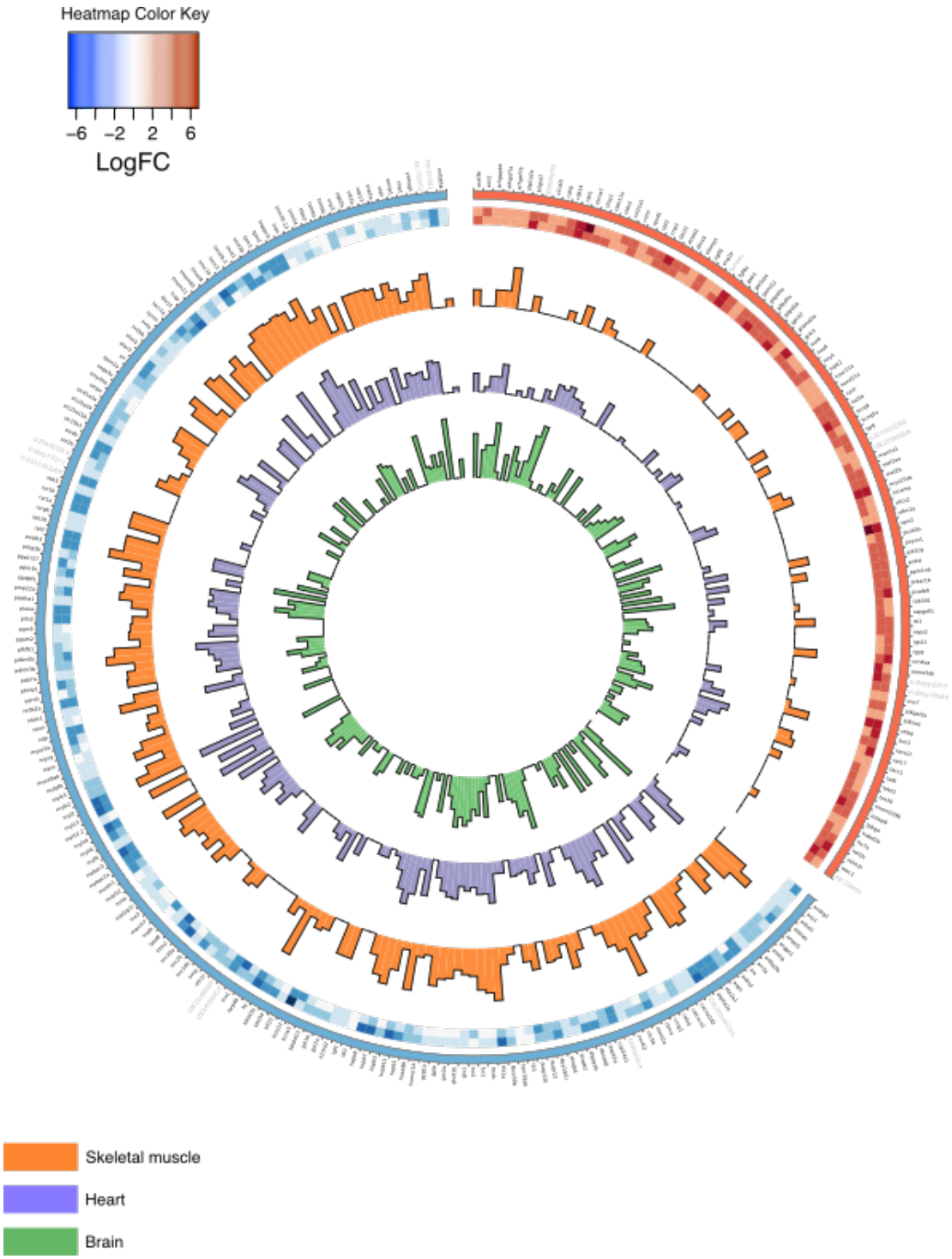
Figure 3



Given the information provided by the category-based functional annotation, and to better understand the functional differences in terms of gene expression between EO and SM,

independently from the species analysed, a literature search was performed on the shared set of differentially expressed genes between *C. compressirostris* and *C. tshokwe.* For each gene, phenotypic information consequent to its mis-expression (e.g., via knockdowns or non-sense mutations) in *D. rerio* was retrieved from the "Zebrafish Model Organism Database" (ZFIN; http://zfin.org/). All shared genes were divided into five functional classes, which synthesize the categories reported in Figure 3b. These functional classes are: "electrical activity" (genes responsible for the differential accumulation and transfer of ions across the plasma membrane), "muscular activity" (genes important for keeping a functional muscle phenotype), "metabolism" (genes involved in metabolic pathways), "transcription factors" (genes regulating gene expression), and "signal transduction" (molecules involved in signalling pathways) (Tables 2-6). Many of the genes present in the class "electrical activity" are up-regulated in the EO (Table 2); they include genes coding for Na+/K+ pumps (*atp1a2a*), voltage-gated sodium (*scn4aa*) and potassium channels (*kcnq5a*), and cholinergic receptors (*chrna7*). However, other voltage-gated ion channels result to be down-regulated in the EO (*kcna3*, *cacna2d2*). There are then two members of the subfamily J of inwardly-rectifying potassium channels that show distinct patterns of expression, with one member (*kcnj9*) up-regulated in EO and the other (*kcnj12*) down-regulated. All the genes included in the class "muscular activity" are down-regulated in the EO (Table 3). As far as the "metabolism" genes are concerned (Table 4), most of the EO up-regulated transcripts are involved in the metabolism of fatty acids, glycerol, and phospholipids (e.g., *acsl3b*, *gdpd4a*, *cds1*), whereas the down-regulated transcripts are more involved in muscle-specific energy production processes like glycolysis (*aldoab*) and gluconeogenesis (*gpib*). Among transcription factors (table 5), two of the four known myogenic factors (transcription factors that activate the expression of sarcomeric proteins), are down-regulated in the EO (*myog*, *myf6*), while the other two (*myoD, myf5*) do not show significant differences in expression between the two tissues. Two basic helix-loop-helix (bHLH) transcription factors (*hey1*, *hes6*) and one co-factor (*her6*) are up-regulated in the electric organ. Two myocyte enhancer factors (*mef2aa*, *mef2b*) show high levels of expression in the EO, whereas two regulators of SM cell proliferation (*six1b*, *six4b*) are poorly expressed in the EO. Most of the EO up-regulated genes involved in signal transduction (Table 6) belong to the G-protein coupled receptor (GPCR) signalling pathway (e.g. *arhgef7a*, *arhgef7b*, *gpr22*) and to the fibroblast growth-factor receptor (FGFR) signalling pathway (e.g. *fgf8a*, *kal1b*)

Figure 4



Discussion

Functional annotation of the 267 differentially expressed genes that are shared between *C. compressirostris* and *C.tshokwe* has revealed marked differences in terms of metabolic pathways, classes of ion channels and regulatory networks, which are probably critical to the

physiological differences between the electric organ and the skeletal muscle. A comparison between the differentially expressed genes identified in this study and the baseline expression levels of human skeletal muscle, heart and brain tissues, obtained from the Illumina body map experiment (http://www.ebi.ac.uk/gxa/experiments/E-MTAB-513) shows that the human orthologs of the EO up-regulated genes are normally more expressed in the brain than in skeletal muscle or heart. This finding seems to suggests that the electric organ might undergo a sort of "neuronalization" process during its differentiation from myogenic tissues possibly due to: i) the complex patterns of innervations that are typical of the electrocytes [21, 22] and ii) the presence of several ion channels, which are normally highly expressed in nervous tissue and that are necessary for increasing electrical conductivity.
The observed differences in terms of gene expression between EO and SM suggest that the metabolic machinery of the electric organ could be mainly devoted to the production and turn-over of membrane structures. This form of specialization might be necessary in order to keep the peculiar anatomy of the momryrids' electrocyte, which is formed by a complex system of membrane ex-vaginations and protrusions (stalks) that strongly influences the electrical properties of the EO [21]. Additionally, a product of phospholipids metabolism, phosphatidylinositol 4,5-bisphosphate, is known for increasing the activity of ion channels [23].

The up-regulation of the *atp1a2a* gene is explained by the fact that its product, the Na^+/K^+ ATP-ase, is fundamental for keeping the electrochemical gradient across the plasma membrane. Voltage-gated ion channels, on the other hand, are important for dissipating the electric potential generated by the ATP-ases and therefore for producing an EOD in response to an action potential. In the electric organ of the analyzed species, one gene coding for a voltage-gated sodium channel (*scn4aa*) is highly expressed in the electric organ, as already observed in other mormyrids and gymnotiform electric fishes [17, 24]. Other over-expressed genes that increase cell excitability are the potassium channels *kcnq5a* and *kcnj9*.

The electric organ of mormyrid fishes still presents non-functional myofibrils; this vestigial presence is reflected in the transcriptome by the moderate expression of a few myosin and actin coding genes. However, most of the genes that are responsible for the assembly and organization of the sarcomere are down-regulated (e.g *tcap*); as well as the genes involved in calcium compartmentalization (e.g. *atp2a1*, *atp2a2*, *casq1a*) and excitation-contraction coupling (e.g. *ryr1*, *stac3*, *jph1*).

The great number of transcription factors (23) and signal transducers (50), identified as differentially expressed, suggests the important role played by regulatory networks in

determining phenotypic differences between EO and SM. For instance the two bHLH transcription factors *hey1* and *hes6*, in co-operation with *her6*, are known to negatively regulate the expression of myogenic factors in several model organisms [25, 26].

Conclusions

The analysis of differentially expressed genes between skeletal muscle and electric organ in two species of African weakly-fishes suggests that: i) the loss of the contractile activity and the decoupling of the excitation-contraction processes are reflected by the down-regulation of the corresponding genes in the electric organ; ii) the metabolic activity of the EO might be specialized towards the production and turn-over of membrane structures; iii) several ion channels are highly expressed in the EO in order to increase excitability; and iv) several myogenic factors might be down-regulated by transcription repressors in the EO.

Methods

Specimen collection

Several specimens of *C. compressirostris* and *C. tshokwe* were collected in the wild during a sampling campaign conducted at the Congo River rapids south of Brazzaville (Republic of the Congo, August/September 2012). For the present study, adult female specimens for each species were selected, kept in captivity for a maximum period of two weeks and then euthanized for tissue sample collection. Gender and sexual maturity were assessed after dissection by checking for the presence of mature ovaries in the selected specimens. Electric organ and skeletal muscle tissue samples were dissected from the caudal peduncle and the posterior trunk musculature respectively (see Fig. 3a) and immediately transferred into RNAlater® (Life Technologies). The research followed internationally recognized guidelines and applicable national legislation. We received ethical approval from the deputy of animal welfare of the University of Potsdam.

RNA extraction and cDNA library preparation

The dissected tissues were processed at the University of Potsdam for RNA extraction: they were first removed from the RNAlater®-containing vials; shock frozen in liquid nitrogen and then homogenized into a buffer containing guanidine isothiocyanate and β-mercaptoethanol using a Mini-beadbeater-1 (Biospec). Total RNA was extracted using the RNeasy® Mini Kit

(Qiagen), RNA quality and concentration was inspected using a Fragment Analyzer™ (Advanced Analytical Technologies, Inc.).

For the present study eight cDNA libraries were selected for sequencing (one library per species per tissue; two biological replicates), each library was obtained by pooling the total RNA from a minimum of two to a maximum of 4 different individuals (see Additional file 1). The paired-end (100nt), strand-specific cDNA libraries were prepared using the NEXTflex™ Directional RNA-Seq Kit V2 (dUTP based) (Bioo Scientific); preparation was performed in six steps: i) mRNA enrichment from total RNA via polyA selection; ii) fragmentation; iii) first and second strand syntheses; iv) A-tailing; v) adapter and barcode ligation and vi) PCR amplification. Fragment size distribution and quality was estimated using an Agilent 2100 Bioanalyzer with the High Sensitivity DNA Chip.

Transcriptome sequencing, assembly and annotation

Transcriptome sequencing was performed at the Max Delbrück Center for Molecular Medicine; the multiplexed cDNA libraries were sequenced using one lane of an Illumina HiSeq2000 sequencing system. After sequencing, the resulting raw reads

were subject to five processing steps using the program Flexbar v2.4 [27] : i) filtering reads with uncalled bases; ii) trimming of reads at 3'-end to get a minimum average Phred quality score of 20; iii) barcode detection, removal and reads separation; iv) adapter detection and removal, and v) filtering of reads shorter than 20 bp after trimming. Quality control of both raw and processed reads was performed with FastQC v0.10.1 (Babraham Bioinformatics).

The processed reads were assembled *de novo* (i.e., without using a reference genome) with Trinity r20131110 [28] (kmer length = 25). Two reference transcriptomes were produced from *C. compressirostris* and *C. tshokwe*, respectively, by assembling together the reads obtained from the EO and SM libraries. Combining all reads across all tissues and all biological replicates for each species into a single RNA-seq dataset allowed to correctly compare transcript abundances from the analysed tissues by aligning the short reads from each library independently onto the same set of reference transcripts (see below)[29].

Transcriptome annotation was conducted using the stand-alone version of the blastx algorithm implemented in Blast+ v2.2.29 [30] (E-value cutoff = 10^{-10}) against the proteome of *Danio rerio* (Uniprot ID = UP000000437). Likely coding sequences were extracted from Trinity transcripts using TransDecoder (http://transdecoder.github.io/) and the longest translated Open Reading Frames (ORFs) were reported (Table 1). Protein domains were searched on the PFAM database (Pfam-A.hmm available at http://pfam.xfam.org/) using HMMER v3.1b1. The retrieved ORFs were later annotated by "blasting" them against the SwissProt (http://web.expasy.org/docs/swiss-prot_guideline.html) database using the blastp algorithm. Transcripts' completeness was assessed by computing the proportion of transcripts and ORFs that matched to full-length top hits in their respective searches using the Perl script "*analyze_blastPlus_topHit_coverage.pl*" (provided with Trinity) (Figure 1 and Additional file 8) [29].

Transcript abundance quantification and DE-analysis.

Short reads from the EO and SM libraries from *C. compressirostris* and *C. tshokwe* were individually mapped to their respective transcriptome assemblies using Bowtie v1.0.0 [31] with default parameters. Gene expression levels were estimated using RSEM v1.2.12 [32]. Putative transcript artifacts and lowly expressed transcripts were filtered out using the Perl script "*filter_fasta_by_rsem_values.pl*" (provided with Trinity).

Differential expression analysis was performed using the Bioconductor package edgeR [33] (minimum fold change = 4, p-value cutoff = 0.001 after FDR correction). The differentially

expressed transcripts were then subject to an enrichment analysis using the Cytoscape plugin ClueGO v2.1.4 [34, 35], in order to identify over-represented functional categories from the Gene Ontology (GO) database [36] (http://geneontology.org), as well as over-represented biological pathways from KEGG (http://www.genome.jp/kegg/) and Reactome (http://www.reactome.org/) [37, 38]. Statistical significance was assessed using a Fisher's exact test with FDR p-value correction (≤0.05).

Data Availability

All the Illumina reads used for this study are available at the Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra), under the accession number SRP050174.

Competing Interests

The authors declare that they have no competing interests.

Authors' contributions

FL participated in specimen sampling, extracted the RNA, carried out the bioinformatic analyses and drafted the manuscript. FK participated in specimen sampling, determined sex and species for the collected samples and participated in manuscript drafting and supervision. IK produced the cDNA libraries and prepared them for sequencing. CD participated in experimental design and provided sequencing resources. RT participated in specimen sampling, conceived and supervised the study. All authors read and provided advice on the manuscript.

Acknowledgements

Funding was provided by the University of Potsdam and the Leibniz-SAW-project GENART. Computational resources were kindly provided by the Unit of Bioinformatics (J. Selbig, S. Hartmann) at the University of Potsdam. The authors are immensely grateful to Dr. Victor Mamonekene (Université Marien Ngouabi, Brazzaville) for the logistic and scientific support provided during their stay in Congo.

References

- Bass A: **Electric organs revisited: evolution of a vertebrate communication and orientation organ**. In *Electroreception*. Edited by Bullock TH, Heiligenberg W. New York: Wiley; 1986:13–70.
- Kirschbaum F: Myogenic electric organ precedes the neurogenic organ in apteronotid fish. *Naturwissenschaften* 1983, **70**:205–207.
- Daget J, Gosse JP, Thys van den Audenaerde DFE: *Check List of the Freshwater Fishes of Africa= Catalogue Des Poissons D'eau Douce d'Afrique. Vol I.* Paris/Tervuren: ORSTOM/MRAC; 1984:410.
- Lissmann HW: **On the function and evolution of electric organs in fish**. *J Exp Biol* 1958, **35**:156–191
- Lissmann HW, Machin KE: The mechanism of object location in *Gymnarchus niloticus* and similar fish. *J Exp Biol* 1958, 35:451–486.
- Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R: **Electrifying love: electric fish use species-specific discharge for mate recognition**. *Biol Lett* 2009, **5**:225–228.
- Kramer B, Kuhn B: **Species recognition by the sequence of discharge intervals in weakly electric fishes of the genus *Campylomormyrus* (Mormyridae, Teleostei)**. *Anim Behav* 1994, **48**:435–445.
- Bratton BO, Kramer B: Patterns of the electric organ discharge during courtship and spawning in the mormyrid fish, *Pollimyrus isidori*. *Behav Ecol Sociobiol* 1989, 24:349–368.
- Crawford JD: Sex recognition by electric cues in a sound-producing mormyrid fish, *Pollimyrus isidori*. *Brain Behav Evol* 1991, 38:20–38.
- Feulner PGD, Kirschbaum F, Mamonekene V, Ketmaier V, Tiedemann R: Adaptive radiation inAfrican weakly electric fish (Teleostei: Mormyridae: *Campylomormyrus*): a combined molecular and morphological approach. *J Evol Biol* 2007, 20:403–414.
- Feulner P, Kirschbaum F, Tiedemann R: Adaptive radiation in the Congo River: An ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; *Campylomormyrus*). *J Physiol* 2008, **102**:340–346.
- Carlson BA, Hasan SM, Hollmann M, Miller DB, Harmon LJ, Arnegard ME: **Brain evolution triggers increased diversification of electric fishes.** *Science* 2011, **332**:583–6.
- Arnegard ME, McIntyre PB, Harmon LJ, Zelditch ML, Crampton WGR, Davis JK, Sullivan JP, Lavoué S, Hopkins CD: **Sexual signal evolution outpaces ecological divergence during electric fish species radiation.** *Am Nat* 2010, **176**:335–56.

- Szabo T: Development of the electric organ of Mormyridae. *Nature* 1960, 188:760–762.
- 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, 11:913–934.
- 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**:1222–1230.
- Gallant JR, Hopkins CD, Deitcher DL: Differential expression of genes and proteins between electric organ and skeletal muscle in the mormyrid electric fish *Brienomyrus brachyistius*. *J Exp Biol* 2012, 215:2479–94.
- Gallant JR, Traeger LL, Volkening JD, Moffett H, Chen PH, Novina CD, Phillips GN, Anand R, Wells GB, Pinch M, Guth R, Unguez GA, Albert JS, Zakon HH, Samanta MP, Sussman MR: Genomic basis for the convergent evolution of electric organs. *Science* 2014, 344:1522–1525.
- Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009, 10:57–63.
- Ekblom R, Galindo J: Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 2011, 107:1–15.
- Moller P: Electric organs. In *Electr Fishes Hist Behav*. Edited by Moller P. London: Chapman & Hall; 1995:385–402.
- Paul C, Mamonekene V, Vater M, Feulner PGD, Engelmann J, Tiedemann R, Kirschbaum F: Comparative histology of the adult electric organ among four species of the genus *Campylomormyrus* (Teleostei: Mormyridae). *J Comp Physiol A* 2015, in press.
- Suh B-C, Hille B: Regulation of ion channels by phosphatidylinositol 4,5-bisphosphate. *Curr Opin Neurobiol* 2005, 15:370–8.
- Zakon HH, Lu Y, Zwickl DJ, Hillis DM: Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. *Proc Natl Acad Sci U S A* 2006, 103:3675–80.
- Fischer A, Gessler M: Delta-Notch--and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. *Nucleic Acids Res* 2007, 35:4583–96.
- Buas MF, Kabak S, Kadesch T: The Notch effector Hey1 associates with myogenic target genes to repress myogenesis. *J Biol Chem* 2010, 285:1249–58.

- Dodt M, Roehr JT, Ahmed R, Dieterich C: FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* 2012, 1:895–905.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, 29:644–52.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A: *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 2013, 8:1494–512.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: BLAST+: architecture and applications. *BMC Bioinformatics* 2009, 10:421.
- Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, 10:R25.
- Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, 12:323.
- Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26:139–40.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, 13:2498–504.
- Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J: **ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.** *Bioinformatics* 2009, **25**:1091–3.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25–9.

- Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic Acids Res* 2000, **28**:27–30.
- Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005, **33**(Database issue):D428–32.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**:1639–45.
- Doganli C, Kjaer-Sorensen K, Knoeckel C, Beck HC, Nyengaard JR, Honoré B, Nissen P, Ribera A, Oxvig C, Lykke-Hartmann K: **The α2Na+/K+-ATPase is critical for skeletal and heart muscle function in zebrafish.** *J Cell Sci* 2012, **125**(Pt 24):6166–75.
- Friedrich T, Lambert AM, Masino MA, Downes GB: **Mutation of zebrafish dihydrolipoamide branched-chain transacylase E2 results in motor dysfunction and models maple syrup urine**
- Chen Y-H, Pai C-W, Huang S-W, Chang S-N, Lin L-Y, Chiang F-T, Lin J-L, Hwang J-J, Tsai C-T: **Inactivation of Myosin binding protein C homolog in zebrafish as a model for human cardiac hypertrophy and diastolic dysfunction.** *J Am Heart Assoc* 2013, **2**:e000231.
- Seguchi O, Takashima S, Yamazaki S, Asakura M, Asano Y, Shintani Y, Wakeno M, Minamino T, Kondo H, Furukawa H, Nakamaru K, Naito A, Takahashi T, Ohtsuka T, Kawakami K, Isomura T, Kitamura S, Tomoike H, Mochizuki N, Kitakaze M: **A cardiac myosin light chain kinase regulates sarcomere assembly in the vertebrate heart.** *J Clin Invest* 2007, **117**:2812–24.
- Postel R, Vakeel P, Topczewski J, Knöll R, Bakkers J: **Zebrafish integrin-linked kinase is required in skeletal muscles for strengthening the integrin-ECM adhesion complex.** *Dev Biol* 2008, **318**:92–101.
- Jurynec MJ, Xia R, Mackrill JJ, Gunther D, Crawford T, Flanigan KM, Abramson JJ, Howard MT, Grunwald DJ: **Selenoprotein N is required for ryanodine receptor calcium release channel activity in human and zebrafish muscle.** *Proc Natl Acad Sci U S A* 2008, **105**:12485–90.
- Dowling JJ, Arbogast S, Hur J, Nelson DD, McEvoy A, Waugh T, Marty I, Lunardi J, Brooks S V, Kuwada JY, Ferreiro A: **Oxidative stress and successful antioxidant treatment in models of RYR1-related myopathy.** *Brain* 2012, **135**:1115–27.

- Li H, Zhong Y, Wang Z, Gao J, Xu J, Chu W, Zhang J, Fang S, Du SJ: **Smyd1b is required for skeletal and cardiac muscle function in zebrafish.** *Mol Biol Cell* 2013, **24**:3511–21.
- Horstick EJ, Linsley JW, Dowling JJ, Hauser MA, McDonald KK, Ashley-Koch A, Saint-Amant L, Satish A, Cui WW, Zhou W, Sprague SM, Stamm DS, Powell CM, Speer MC, Franzini-Armstrong C, Hirata H, Kuwada JY: **Stac3 is a component of the excitation-contraction coupling machinery and mutated in Native American myopathy.** *Nat Commun* 2013, **4**:1952.
- Zhang R, Yang J, Zhu J, Xu X: **Depletion of zebrafish Tcap leads to muscular dystrophy via disrupting sarcomere-membrane interaction, not sarcomere assembly.** *Hum Mol Genet* 2009, **18**:4130–40.
- Pan W, Pham VN, Stratman AN, Castranova D, Kamei M, Kidd KR, Lo BD, Shaw KM, Torres-Vazquez J, Mikelis CM, Gutkind JS, Davis GE, Weinstein BM: **CDP-diacylglycerol synthetase-controlled phosphoinositide availability limits VEGFA signaling and vascular morphogenesis.** *Blood* 2012, **120**:489–98.
- Molt S, Bührdel JB, Yakovlev S, Schein P, Orfanos Z, Kirfel G, Winter L, Wiche G, van der Ven PFM, Rottbauer W, Just S, Belkin AM, Fürst DO: **Aciculin interacts with filamin C and Xin and is essential for myofibril assembly, remodeling and maintenance.** *J Cell Sci* 2014, **127**(Pt 16):3578–92
- Bitomsky N, Conrad E, Moritz C, Polonio-Vallon T, Sombroek D, Schultheiss K, Glas C, Greiner V, Herbel C, Mantovani F, del Sal G, Peri F, Hofmann TG: **Autophosphorylation and Pin1 binding coordinate DNA damage-induced HIPK2 activation and cell death.** *Proc Natl Acad Sci U S A* 2013, **110**:E4203–12.
- Wang Y, Qian L, Dong Y, Jiang Q, Gui Y, Zhong TP, Song H: **Myocyte-specific enhancer factor 2A is essential for zebrafish posterior somite development.** *Mech Dev* 2006, **123**:783–91.
- Gyda M, Wolman M, Lorent K, Granato M: **The tumor suppressor gene retinoblastoma-1 is required for retinotectal development and visual function in zebrafish.** *PLoS Genet* 2012, **8**:e1003106.
- Filippi A, Jainok C, Driever W: **Analysis of transcriptional codes for zebrafish dopaminergic neurons reveals essential functions of Arx and Isl1 in prethalamic dopaminergic neuron development.** *Dev Biol* 2012, **369**:133–49.
- Wang Y-H, Li C-K, Lee G-H, Tsay H-J, Tsai H-J, Chen Y-H: **Inactivation of zebrafish mrf4 leads to myofibril misalignment and motor axon growth disorganization.** *Dev Dyn* 2008, **237**:1043–50.

- Hinits Y, Osborn DPS, Hughes SM: **Differential requirements for myogenic regulatory factors distinguish medial and lateral somitic, cranial and fin muscle fibre populations.** *Development* 2009, **136**:403–14.
- Liu Y, Semina E V: pitx2 Deficiency results in abnormal ocular and craniofacial development in zebrafish. *PLoS One* 2012, 7:e30896.
- Bessarab DA, Chong S-W, Srinivas BP, Korzh V: Six1a is required for the onset of fast muscle differentiation in zebrafish. *Dev Biol* 2008, 323:216–28.
- Bontems F, Fish RJ, Borlat I, Lembo F, Chocu S, Chalmel F, Borg J-P, Pineau C, Neerman-Arbez M, Bairoch A, Lane L: C2orf62 and TTC17 are involved in actin organization and ciliogenesis in zebrafish and human. *PLoS One* 2014, 9:e86476.
- Albertson RC, Yelick PC: Fgf8 haploinsufficiency results in distinct craniofacial defects in adult zebrafish. *Dev Biol* 2007, 306:505–15.
- Lee J-A, Anholt RRH, Cole GJ: Olfactomedin-2 mediates development of the anterior central nervous system and head structures in zebrafish. *Mech Dev* 2008, 125:167–81.
- Lamont RE, Vu W, Carter AD, Serluca FC, MacRae CA, Childs SJ: Hedgehog signaling via angiopoietin1 is required for developmental vascular stability. *Mech Dev* 2010, 127:159–68.
- Xie H, Fan X, Tang X, Wan Y: The LIM Protein fhlA is Essential for Heart Chamber Development in Zebrafish Embryos. *Curr Mol Med* 2013, 13:979–992.
- Ruparelia A a, Zhao M, Currie PD, Bryson-Richardson RJ: Characterization and investigation of zebrafish models of filamin-related myofibrillar myopathy. *Hum Mol Genet* 2012, 21:4073–83.
- Lin C-Y, Chen J-S, Loo M-R, Hsiao C-C, Chang W-Y, Tsai H-J: MicroRNA-3906 regulates fast muscle differentiation through modulating the target gene homer-1b in zebrafish embryos. *PLoS One* 2013, 8:e70187.
- Gupta VA, Ravenscroft G, Shaheen R, Todd EJ, Swanson LC, Shiina M, Ogata K, Hsu C, Clarke NF, Darras BT, Farrar MA, Hashem A, Manton ND, Muntoni F, North KN, Sandaradura SA, Nishino I, Hayashi YK, Sewry CA, Thompson EM, Yau KS, Brownstein CA, Yu TW, Allcock RJN, Davis MR, Wallgren-Pettersson C, Matsumoto N, Alkuraya FS, Laing NG, Beggs AH: Identification of KLHL41 mutations implicates BTB-Kelch-mediated ubiquitination as an alternate pathway to myofibrillar disruption in Nemaline myopathy. *Am J Hum Genet* 2013, 93:1108–17.
- Özhan G, Sezgin E, Wehner D, Pfister AS, Kühl SJ, Kagermeier-Schenk B, Kühl M, Schwille P, Weidinger G: Lypd6 enhances Wnt/β-catenin signaling by promoting Lrp6 phosphorylation in raft plasma membrane domains. *Dev Cell* 2013, 26:331–45.

- Jiang Z, Song J, Qi F, Xiao A, An X, Liu N, Zhu Z, Zhang B, Lin S: Exdpf is a key regulator of exocrine pancreas development controlled by retinoic acid and ptf1a in zebrafish. *PLoS Biol* 2008, 6:e293.
- Komoike Y, Fujii K, Nishimura A, Hiraki Y, Hayashidani M, Shimojima K, Nishizawa T, Higashi K, Yasukawa K, Saitsu H, Miyake N, Mizuguchi T, Matsumoto N, Osawa M, Kohno Y, Higashinakagawa T, Yamamoto T: Zebrafish gene knockdowns imply roles for human YWHAG in infantile spasms and cardiomegaly. *Genesis* 2010, 48:979–992.

Figures

Figure 1 - Distribution of length coverage between *Campylomormyrus* Trinity transcripts and corresponding top-blast hits (*D. rerio* proteome). Histogram showing the distribution of the percent in length of the sequences in the *D. rerio* proteome that aligns to the *Campylomormyrus* Trinity contigs. Numbers on the x-axis indicate the upper limit of the binned interval (e.g. 100 is the upper value of the interval 100-91).

Figure 2 - Number of differentially expressed genes.

Venn diagram showing the amount of differentially expressed genes within each *Campylomormyrus* species' transcriptome (full circles) and the amount of differentially expressed genes shared between the two *Campylomormyrus* species (overlapping area). The amount of genes that are up (EO[+])- or down (EO[-])-regulated in the electric organ are reported for each dataset.

Figure 3 - Illustration of *C. compressirostris* and *C. tshokwe* and functional annotation results.

a. Pictures illustrating *C. compressirostris* and *C. tshokwe*. The red circle indicate the localization (and site of excision) of the electric organ (EO); the blue oval indicates the localization (and site of excision) of the deep lateral muscle.

b. Pie charts showing the composition in terms of enriched functional categories (GO) and pathways (KEGG, Reactome) for each cluster of differentially expressed genes. EO[+] = up-regulated in the electric organ; EO [-] = down-regulated in the electric organ.

Figure 4 - Comparison of the differentially expressed genes between EO and SM with the expression levels of their human orthologs in skeletal muscle, heart and brain.

Circos graph [39] showing the set of differentially expressed genes common to *C. compressirostris* (outer heatmap) and *C. tshokwe* (inner heatmap). The three histograms indicate the baseline expression levels of the human orthologs in skeletal muscle, heart and brain tissues. The two sets of genes are separated in EO

down-regulated (blue) and EO up-regulated (red). Light-grey labels indicate genes without corresponding human orthologs.

Tables

Table 1 - Assembly statistics

	C. compressirostris	*C. tshokwe*
Trinity contigs	260,598	369,030
# of retrieved ORFs	139,963	228,306
# of unique PFAM domains	9,986	9,941
# of contigs matching *D. rerio* proteome	108,705	159,263
# of unique hits to *D. rerio* proteome	18,458	19,363
N50	3,010	3,597
Average contig length	1,392.16	1,672.28
Total assembled bases	362,794,286	617,123,151

Table 2 - Differentially expressed genes in the functional class "Electrical activity"

For each of the shared differentially expressed gene are reported: the gene and protein names obtained from the top hit blast results against the proteome of *D. rerio*; whether it is up(+)- or down(-)- regulated in the EO; its function or pathway

(or both when available); the phenotypic effect on *D. rerio* of its mis-expression (when available).

Gene	Protein Name	Expression in EO	Pathway/Function	Disrupted Phenotype	Reference
atp1a2a	ATPase, Na+/K+ transporting, alpha 2a polypeptide	+	Ion channel transport	Impaired depolarization of the resting membrane potential in slow-twitch fibers of skeletal muscles.	[40]
chrna7	cholinergic receptor, nicotinic, alpha 7 (neuronal)	+	Activation of Nicotinic Acetylcholine Receptors		
kcnj9	potassium inwardly-rectifying channel, subfamily J, member 9	+	Potassium Channels; GABA receptor activation		
kcnq5a	potassium voltage-gated channel, KQT-like subfamily, member 5a	+	Potassium Channels; Synaptic transmission ion currents		
grik3	Glutamate Receptor, Ionotropic, Kainate 3	+	Transmission across Chemical Synapses		
scn4aa	sodium channel, voltage-gated, type IV, alpha, a	+	Ion channel transport; Axon guidance		
kcna3	potassium voltage-gated channel, shaker-related subfamily, member 3	-	Potassium Channels; Transmission across Chemical Synapses		
kcnj12	potassium inwardly-rectifying channel, subfamily J, member 12	-	Potassium Channels; GABA receptor activation		
cacna2d2	calcium channel, voltage-dependent, alpha 2/delta subunit 2	-	Ion channel transport		

Table 3 - Differentially expressed genes in the functional class "Muscular activity"

For each of the shared differentially expressed gene are reported: the gene and protein names obtained from the top hit blast results against the proteome of D. rerio; whether it is up(+)- or down(-)- regulated in the EO; its function or pathway (or both when available); the phenotypic effect on D. rerio of its mis-expression (when available).

Gene	Protein Name	Expression in EO	Pathway/Function	Reference
atp2a1	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	-	Muscle contraction	[41]
atp2a2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2a	-	regulation of heart contraction	[42]
casq1a	calsequestrin 1a	-	Calcium homeostasis	
jph1a	junctophilin 1a	-	structural constituent of muscle	
jph1b	junctophilin 1b	-	structural constituent of muscle	
myl2a	myosin, light chain 2a, regulatory, cardiac, slow	-	Striated Muscle Contraction	
mybpc2a	myosin binding protein C, fast type a	-	Striated Muscle Contraction	
mybpc3	myosin binding protein C, cardiac	-	Cardiac muscle contraction	[43]
myhb	myosin, heavy chain b	-	Striated Muscle Contraction	
myl10	myosin, light chain 10, regulatory	-	Regulation of actin	

			cytoskeleton; Focal adhesion	
myl12.2	myosin, light chain 12, genome duplicate 2	-	Striated Muscle Contraction	
mylk2	myosin light chain kinase 2	-	Focal adhesion; Regulation of actin cytoskeleton	
mylk3	myosin light chain kinase 3	-	Focal adhesion; Regulation of actin cytoskeleton	[44]
mylpfb	myosin light chain, phosphorylatable, fast skeletal muscle b	-	Focal adhesion; Regulation of actin cytoskeleton	
myo18ab	myosin XVIIIAb	-	Signaling by FGFR	
myoz3a	myozenin 3a	-	Calcineurin signaling	
nexn	nexilin (F actin binding protein)	-	cardiac muscle fiber development	
parvb	parvin, beta	-	Focal adhesion; Cell junction organization	[45]
pdlim3b	PDZ and LIM domain 3b	-		
pdlim5b	PDZ and LIM domain 5b	-		
pvalb3	parvalbumin 3	-	calcium ion homeostasis	
ryr1a	ryanodine receptor 1a (skeletal)	-	calcium ion channel transport	[46]
ryr1b	ryanodine receptor 1b (skeletal)	-	calcium ion channel transport	[47]
tnnc2	troponin C type 2 (fast)	-	Striated Muscle Contraction	
smpx	small muscle protein, X-linked	-	Striated Muscle Contraction	
smyd1b	SET and MYND domain containing 1b	-	Muscle Development	[48]
srl	sarcalumenin	-	calcium ion homeostasis	
stac3	SH3 and cysteine rich domain 3	-	Striated Muscle Contraction	[49]
tcap	titin-cap (telethonin)	-	Striated Muscle Contraction	[50]
tmod4	tropomodulin 4 (muscle)	-	Muscle contraction	
tnnc1b	troponin C type 1b (slow)	-	Muscle contraction	

tnni2b.2	troponin I type 2b (skeletal, fast), tandem duplicate 2	-	Striated Muscle Contraction
tnnt1	troponin T type 1 (skeletal, slow)	-	Muscle contraction
tnnt3b	troponin T type 3b (skeletal, fast)	-	Striated Muscle Contraction
tpm1	tropomyosin 1 (alpha)	-	Striated Muscle Contraction
tpm2	tropomyosin 2 (beta)	-	Striated Muscle Contraction
trdn	triadin	-	Muscle contraction
trim54	tripartite motif containing 54	-	Titin-kinase regulation
xirp1	xin actin-binding repeat containing 1	-	
myl13	myosin, light chain 13	-	

Table 4 - Differentially expressed genes in the functional class "Metabolism"

For each of the shared differentially expressed gene are reported: the gene and protein names obtained from the top hit blast results against the proteome of D. rerio; whether it is up(+)- or down(-)- regulated in the EO; its function or pathway (or both when available); the phenotypic effect on D. rerio of its mis-expression (when available).

Gene	Protein Name	Expression in EO	Pathway/Function	Disrupted Phenotype	Reference
acsbg2	acyl-CoA synthetase bubblegum family member 2	-	Fatty acid metabolism		

acsl3b	acyl-CoA synthetase long-chain family member 3b	+	Fatty acid metabolism		
acy1	Aminoacylase-1	-	Aminoacids metabolism		
adssl1	adenylosuccinate synthase like 1	-	Purine metabolism		
aldoab	Fructose-bisphosphate aldolase	-	Glycolysis		
ampd1	Adenosine monophosphate deaminase 1 (Isoform M)	-	Purine metabolism		
aoc2	amine oxidase, copper containing 2	+	beta-Alanine metabolism		
cds1	CDP-Diacylglycerol Synthase 1	+	Glycerophospholipid biosynthesis	Imperfect angiogenesis	[51]
ckma	creatine kinase, muscle a	-	Metabolism of amino acids and derivatives		
ckmt2a	creatine kinase, mitochondrial 2a (sarcomeric)	-	Metabolism of amino acids and derivatives		
cox4i2	cytochrome c oxidase subunit IV isoform 2	-	Oxidative phosphorylation		
cpt2	carnitine palmitoyltransferase 2	+	Fatty acid beta-oxidation		
cyp24a1	cytochrome P450, family 24, subfamily A, polypeptide 1	-	Steroid biosynthesis		
dhrs9	dehydrogenase/reductase (SDR family) member 9	+	Retinol metabolism		
gdpd4a	glycerophosphodiester phosphodiesterase domain containing 4a	+	Glycerol metabolism		
gdpd5a	glycerophosphodiester phosphodiesterase domain containing 5a	+	Glycerol metabolism		
gdpd5b	glycerophosphodiester phosphodiesterase domain containing 5b	+	Glycerol metabolism		
glo1	Glyoxalase 1	-	Pyruvate metabolism		
glud1b	glutamate dehydrogenase 1b	-	Nitrogen metabolism		
got2a	glutamic-oxaloacetic	-	Glucose metabolism;		

	transaminase 2a, mitochondrial		aminoacids metabolism		
gpib	glucose-6-phosphate isomerase b	-	Gluconeogenesis		
idi1	isopentenyl-diphosphate delta isomerase 1	-	Cholesterol biosynthesis		
man1a1	mannosidase, alpha, class 1A, member 1	+	N-Glycan biosynthesis		
me3	malic enzyme 1, NADP(+)-dependent, cytosolic	-	Pyruvate metabolism		
pcyox1	prenylcysteine oxidase 1	-	Terpenoid backbone biosynthesis		
pfkfb1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	-	Glycolysis		
pgam2	phosphoglycerate mutase 2 (muscle)	-	Glycolysis and Gluconeogenesis		
pgm5	Phosphoglucomutase 5	-	Glucuronidation	Failure in myofibril assembly	[52]
ucp3	uncoupling protein 3	-	Respiratory electron transport		
ugp2b	UDP-glucose pyrophosphorylase 2b	-	Glucose metabolism		
gyg1a	glycogenin 1a	-	Glycogen Metabolism		
mid1ip1l	MID1 interacting protein 1, like	-	lipid metabolic process		

Table 5 - Differentially expressed genes in the functional class "Transcription factors"

For each of the shared differentially expressed gene are reported: the gene and protein names obtained from the top hit blast results against the proteome of D.

rerio; whether it is up(+)- or down(-)- regulated in the EO; its function or pathway (or both when available); the phenotypic effect on D. rerio of its mis-expression (when available).

Gene	Protein Name	Expression in EO	Pathway/Function	Disrupted Phenotype	Reference
eng1b	engrailed homeobox 1b	+	neuron fate commitment		
her6	hairy-related 6	+	Notch signaling pathway		
hes6	hes family bHLH transcription factor 6	+	Notch signaling pathway		
hey1	hes-related family bHLH transcription factor with YRPW motif 1	+	Notch signaling pathway		
hipk2	homeodomain interacting protein kinase 2	+	Wnt signaling pathway; p53 Signaling; ERK Signaling	Induced apoptosis	[58]
hoxc11a	homeobox C11a	+			
hoxd11a	homeobox D11a	+			
mef2aa	myocyte enhancer factor 2aa	+	Signaling by FGFR	Abnormal development of posterior somites	[29]
mef2b	myocyte enhancer factor 2b	+	miRs in Muscle Cell Differentiation		
rb1	retinoblastoma 1	+	E2F mediated regulation of DNA replication; Cell cycle	Abnormal retina development	[59]
taf6	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor	+	GPCR Pathway		
arxa	aristaless related homeobox a	-	Axon guidance	Abnormal dopaminergic neurons development	[60]
klf15	Kruppel-like factor 15	-	Adipogenesis		

pbxip1	pre-B-cell leukemia homeobox interacting protein 1	-			
myf6	myogenic factor 6	-	Myogenesis	Disrupted myogenesis	[28]
myog	myogenin	-	Myogenesis	Disrupted myogenesis	[27]
nfatc1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	-	Wnt signaling pathway		
nr0b2a	nuclear receptor subfamily 0, group B, member 2a	-	Nuclear Receptor transcription pathway; NOD-like Receptor Signaling Pathways		
pitx2	paired-like homeodomain 2	-	retinoic acid receptor signaling pathway	Abnormal eye and craniofacial development	[61]
rxrgb	retinoid X receptor, gamma b	-	steroid hormone receptor activity; retinoic acid receptor signaling pathway		
six1b	SIX homeobox 1b	-	regulation of skeletal muscle cell proliferation	Abnormal trunk musculature development	[62]
six4b	SIX homeobox 4b	-	regulation of skeletal muscle cell proliferation		
tbx15	T-box 15	-	regulation of transcription, DNA-templated		

Table 6 - Differentially expressed genes in the functional class "Signal transduction"

For each of the shared differentially expressed gene are reported: the gene and protein names obtained from the top hit blast results against the proteome of *D. rerio*; whether it is up(+)- or down(-)- regulated in the EO; its function or pathway (or both when available); the phenotypic effect on *D. rerio* of its mis-expression (when available).

Gene	Protein Name	Expression in EO	Pathway	Disrupted Phenotype	Reference
arhgef7a	Rho guanine nucleotide exchange factor (GEF) 7a	+	Regulation of actin cytoskeleton; Signaling by GPCR; Signaling by FGFR		
arhgef7b	Rho guanine nucleotide exchange factor (GEF) 7b	+	Regulation of actin cytoskeleton; Signaling by GPCR; Signaling by FGFR		
catip	ciliogenesis associated TTC17 interacting protein	+	actin filament polymerization	Imperfect ciliogenesis	Bontems et al., 2014
fgf8a	fibroblast growth factor 8a	+	Signaling by FGFR; Regulation of actin cytoskeleton	Imperfect morphogenesis	Albertson et al., 2007
kal1b	Kallmann syndrome 1b sequence	+	Signaling by FGFR		
gpr22	G protein-coupled receptor 22	+	G-protein coupled receptor signaling pathway		
reps2	RALBP1 associated Eps domain containing 2	+	EGFR1 Signaling Pathway		
olfcs2	olfactory receptor C family, s2	+	G-protein coupled receptor signaling pathway		
olfm2a	olfactomedin 2a	+		Central nervous system development	Lee et al., 2008
opn3	Opsin 3	+	G-protein coupled receptor signaling pathway		
pcsk5b	proprotein convertase subtilisin/kexin type 5b	+	Signaling by FGFR; Signaling by GPCR; NGF processing		
pik3cg	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit gamma	+	Toll-like receptor signaling pathway		

prkar1b	protein kinase, cAMP-dependent, regulatory, type I, beta	+	G-protein coupled receptor signaling pathway		
rgs11	regulator of G-protein signaling 11	+	G-protein coupled receptor signaling pathway		
arhgap44	Rho GTPase activating protein 44	+	G-protein coupled receptor signaling pathway		
rapgefl1	Rap guanine nucleotide exchange factor (GEF)-like 1	+	G-protein coupled receptor signaling pathway		
gab3	GRB2-Associated Binding Protein 3	+	Akt Signaling Pathway		
tpbga	trophoblast glycoprotein a	+	negative regulation of canonical Wnt signaling pathway		
trabd2b	TraB domain containing 2B	+	Wnt signaling pathway		
twf1b	twinfilin actin-binding protein 1b	+	negative regulation of actin filament polymerization		
wwc1	WW and C2 domain containing 1	+	G-protein coupled receptor signaling pathway		
cdk14	cyclin-dependent kinase 14	+	Transcriptional misregulation in cancer		
angpt1	angiopoietin 1	-	ERK Signaling; Akt Signaling; TGF-Beta Pathway; Hedgehog signaling	Imperfect angiogenesis	
asb10	ankyrin repeat and SOCS box containing 10	-	Class I MHC mediated antigen processing and presentation		
calcoco1	calcium binding and coiled-coil domain 1	-	Wnt signaling pathway		
ccng1	cyclin G1	-	p53 signaling pathway; G-protein coupled receptor signaling pathway		
dapk2a	death-associated protein kinase 2a	-	Regulation of Apoptosis		
dusp22b	dual specificity phosphatase 22b	-	TGF-Beta Pathway		
fhl1a	four and a half LIM domains 1a	-	Delta-Notch Signaling Pathway	Abnormal cardiac function	Xie et al., 2013
flncb	filamin C, gamma b (actin binding protein 280)	-	MAPK signaling pathway	Myofibril disruption	Ruparelia et al., 2012
homer1b	homer homolog 1b (Drosophila)	-	FoxO signaling pathway; Regulation of calcium homeostasis	defective phenotypes in fast muscle	Lin et al., 2013
igf1	insulin-like growth factor 1	-	Development IGF 1 receptor signaling; G-protein coupled receptor signaling pathway		

il13ra2	interleukin 13 receptor, alpha 2	-	Akt Signaling; TGF-Beta Pathway; ERK Signaling		
klhl41b	kelch-like family member 41b	-	Regulation of myoblast differentiation	myofibrillar disorganization	Gupta et al., 2013
lnx1	ligand of numb-protein X 1	-	Notch signaling pathway		
lypd6	LY6/PLAUR domain containing 6	-	positive regulation of canonical Wnt signaling pathway	caudal fin decreased size; trunk decreased size	Özhan et al., 2013
myoc	myocilin	-	Wnt signaling pathway		
ndp	Norrie disease (pseudoglioma)	-	Wnt signaling pathway		
pde7a	phosphodiesterase 7A	-	G-protein coupled receptor signaling pathway		
pmp22a	peripheral myelin protein 22a	-	Neural Crest Differentiation		
ppdpfa	pancreatic progenitor cell differentiation and proliferation factor a	-	Negative regulation of RA signaling pathway	Abnormal pancreas development	Jiang et al., 2008
prkg1b	protein kinase, cGMP-dependent, type Ib	-	beta-catenin independent WNT signaling		
sbk3	SH3 domain binding kinase family, member 3	-	MAPK signaling pathway		
mras	muscle RAS oncogene homolog	-	MAPK signaling pathway; G-protein coupled receptor signaling pathway		
plekha1	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1	-	Class I PI3K signaling events		
spon2a	spondin 2a, extracellular matrix protein	-	Integrin Pathway; ERK Signaling		
tacr1a	tachykinin receptor 1a	-	G-protein coupled receptor signaling pathway		
txlnba	taxilin beta a	-	TNF-alpha/NF-kB Signaling Pathway		
txlnbb	taxilin beta b	-	TNF-alpha/NF-kB Signaling Pathway		
ywhag1	3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide 1	-	Cell cycle	Reduced brain size; Increased heart tube diameter	Komoike et al., 2010

Additional files

Additional file 1 – Sequencing statistics

For each of the eight produced libraries we report: the number of pooled individuals, the number of pre- and post- processing read pairs, and the percent of retained reads after quality filtering.

Additional file 2 – Blast results of *C. compressirostris* transcriptome assembly

Tabular blast report showing the results of the blastx comparison between the transcriptome of *C. compressirostris* and the proteome of *D. rerio*.

Additional file 3 – Blast results of *C. tshokwe* transcriptome assembly

Tabular blast report showing the results of the blastx comparison between the transcriptome of *C. tshokwe* and the proteome of *D. rerio*.

Additional file 4 – Changes in gene expression between EO and SM for *C. compressirostris* transcriptome

Results of the edgeR DE-analysis. The table reports the fold-change in the expression levels and relative uncorrected and corrected p-values between EO and SM for all genes in the *C. compressirostris* transcriptome.

Additional file 5 – Changes in gene expression between EO and SM for *C. tshokwe* transcriptome

Results of the edgeR DE-analysis. The table reports the fold-change in the expression levels and relative uncorrected and corrected p-values between EO and SM for all genes in the *C. tshokwe* transcriptome.

Additional file 6 – Transcripts' length distribution of the *C. compressirostris* Trinity assembly.

Additional file 7 – Transcripts' length distribution of the *C. tshokwe* Trinity assembly.

Additional file 8 – Distribution of length coverage between retrieved ORFs and reference database (SwissProt).

7 Article III

Coalescent-based Species Tree Estimation and Delimitation in a Group of Sympatric Weakly-Electric Fish (Osteoglossiphormes, Mormyridae, *Campylomormyrus*).

RUNNING HEAD: SPECIES TREE ESTIMATION IN WEAKLY-ELECTRIC FISH

Coalescent-based Species Tree Estimation and Delimitation in a Group of Sympatric Weakly-Electric Fish (Osteoglossiformes, Mormyridae, *Campylomormyrus*)

FRANCESCO LAMANNA[1], FRANK KIRSCHBAUM[2], ANJA R. R. SCHNEIDER[1], PHILINE G. D. FEULNER[1,3], CHRISTIANE PAUL[1], VICTOR MAMONEKENE[4] AND RALPH TIEDEMANN[1*]

[1] *Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, D-14476, Potsdam, Germany*

[2]*Department of Crop and Animal Sciences, Faculty of Horticulture and Agriculture, Humboldt University, Philippstrasse 13, D-10115 Berlin, Germany*

[3]*current address: Max Planck Institute for Evolutionary Biology, August-Thienemann-Strasse 2, D-24306, Plön, Germany*

Institut de Développement Rural, Université Marien Ngouabi,B.P. 69 Brazzaville, Republic of Congo

[*]*Correspondence to be sent to: University of Potsdam, Karl-Liebknecht-Strasse 24-25, D-14476, Potsdam, Germany; e-mail: tiedeman@uni-potsdam.de.*

Abstract

Understanding the modalities of speciation constitutes one of the principal aims of evolutionary biology. Preliminary steps towards the accomplishment of this task include: i) inference of reliable and robust phylogenetic trees that can account for the actual relationships among the analysed taxa; ii) identification of "independent evolutionary lineages"; iii) selection of the diversification model that fits the data best. In this study, we applied coalescent-based methods in order to: infer a species tree, identify species boundaries and select models of differentiation in a genus of sympatric African weakly-electric fish (Teleostei, Mormyridae, *Campylomormyrus*). Members of this genus make use of Electric Organ Discharges (EODs) for conspecific/mate recognition, a feature which is putatively involved in the onset of prezygotic reproductive isolation in sympatry. We analyzed sequence data collected from 5 molecular markers (one mitochondrial and four nuclear genes) in order to build a species tree of the genus and to delimit species using coalescent-based methods. We additionally used 16 microsatellite loci as an independent line of evidence for the detection of reproductively isolated groups within *Campylomormyrus*. Finally, we applied Approximate Bayesian Computation (ABC) and Posterior Predictive Simulations (PPS) on both datasets to select the diversification model that fitted the data best and to check for allelic introgression among species. [electric fish; multispecies coalescent; species delimitation; approximate bayesian computation; mormyridae]

The study of speciation aims at identifying the relative contributions of the forces - i.e. natural selection, gene flow, genetic drift, etc. - that have shaped the current observable patterns of diversity among taxa. The typical work-flow for the inference of such diversification modalities would involve the following steps: i) estimation of the species tree using molecular markers; ii) determination of species boundaries among the analysed taxonomic units; iii) selection of the diversification model that best fits the observed data. Until recently, it has been common practice to extrapolate species trees from gene trees obtained using one or few concatenated loci, sampled from single individuals per species. This method, however, does not take into account the possible discordances among gene trees, which may arise due to phenomena such as incomplete lineage sorting, occurrence of gene flow among speciating units, hybridization, gene duplication/extinction and may hence possibly lead to the estimation of incorrect species trees (Tajima 1983; Pamilo and Nei 1988; Degnan and Rosenberg 2006, 2009; Kubatko and Degnan 2007). The recent introduction of multilocus coalescent models (Wilson and Balding 1998; Rannala and Yang 2003; Edwards et al. 2007; Kubatko et al. 2009; Heled and Drummond 2010) has determined an important conceptual advance towards the reconciliation of gene and species trees. These models bridge together classic phylogenetic inference with population genetics, by considering intra- and interspecifc diversity and by explicitly modelling the stochastic processes that generate gene trees discordance.

Concurrently, species delimitation methods have undergone a similar progress, alongside with the development of richer and more inclusive species concepts (De Queiroz

2011; Hausdorf 2011). Indeed, the last few years have witnessed a gradual shift from the use of criteria based on rather arbitrary threshold values to define species boundaries - e.g. differences in FST estimates or ratios between intraspecific and interspecific diversity-towards the application of more rigorous, coalescent-based, approaches (O'Meara et al. 2006; Pons et al. 2006; Yang and Rannala 2010; Fujita et al. 2012; Camargo and Sites 2013). The joint application of species tree estimation and species delimitation methods under a coalescent framework is becoming extensively used in systematics, particularly in the study of recently diverged species (Leaché and Fujita 2010; Camargo et al. 2012; Ruane et al. 2013; Satler et al. 2013).

Most of the studies conducted so far, investigated allopatric taxonomic units - species, sub-species, populations -, using geographic information to determine *a priori* the groups to be tested. More difficult is to define prior groupings when the analyzed taxa occur in sympatry, especially in the lack of clear morphological differentiation. In such cases one solution would be to discriminate groups relying on features that are putatively involved in the onset of reproductive isolation - e.g.,

exploitation of different niches or hosts, display of divergent sexual characters, differences in communication systems - and subsequently testing whether the identified units correspond to actual "independent evolving lineages" (Simpson 1951; de Queiroz 1998). In particular, changes in mate recognition systems are known to promote speciation among sympatric populations (Higgie et al. 2000; Jiggins et al. 2001; Barluenga et al. 2006), by increasing assortative mating and, hence, reproductive isolation (Felsenstein 1981; Via 2001; Butlin et al. 2012). One taxonomic group for which differences in mate recognition systems appear to play a central role in reproductive isolation is the family of African weakly-electric fish Mormyridae (Lavoué et al. 2008).

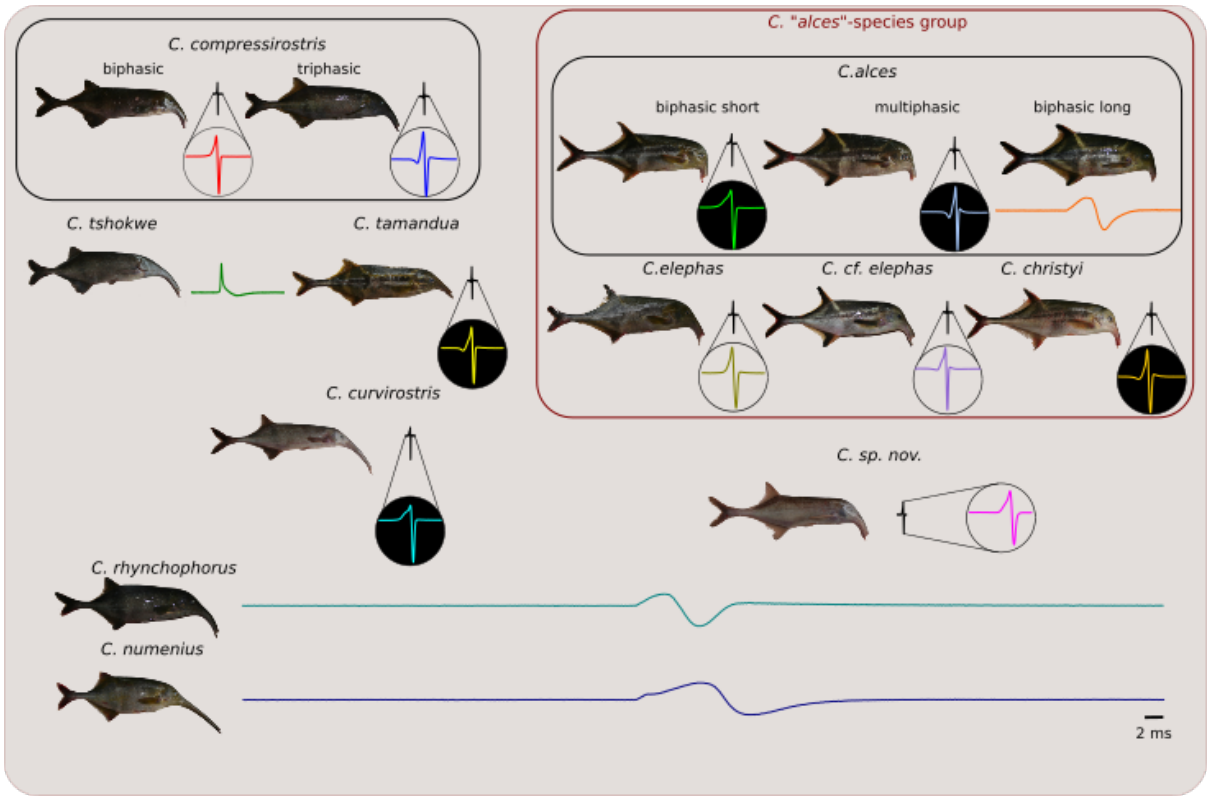
The family Mormyridae contains approximately 200 species, all endemic to African riverine systems (Daget et al. 1991). It constitutes the single largest known assemblage of electric fish and one of the most diverse clades of African freshwater fish (Alves-Gomes and Hopkins 1997). Together with their monospecific sister family Gymnarchidae, they share the presence of an electric organ which allows them to produce weak electric discharges (Lissmann 1951, 1958). Most mormyrids are nocturnal, therefore they use their electric sense to actively locate objects and food in the darkness (Lissmann and Machin 1958; Bastian 1994; von der Emde 1999), and for conspecific/mate recognition. Indeed, mormyrids are able to discriminate between conspecifics and heterospecifics Electric Organ Discharges (EODs), by perceiving differences in waveform shapes and pulse durations (Bratton and Kramer 1989; Crawford 1991; Bullock et al. 2005; Lamml and Kramer 2006).

A taxon of mormyrids for which EOD was demonstrated to play a key role in mate recognition and pair formation is the genus *Campylomormyrus* (Kramer and Kuhn 1994; Feulner et al. 2009a, 2009b). Members of the genus *Campylomormyrus* are characterized by the presence of an elongated trunk-like snout, which may vary dramatically in length and shape across different species (Fig. 1); ecological studies highlighted the link between different snout morphologies and differential feeding habits (Marrero and Winemiller 1993). Moreover, geometric morphometrics analyses showed that most of the morphological variance among different species is confined within the snout region (Feulner et al. 2007, 2008). Both lines of evidence invoke ecological adaptation as the main factor that might have prompted speciation within this genus. Alongside of morphological divergence,

Campylomormyrus species display very remarkable differences in terms of EOD shape and duration (Fig. 1). All members of the genus except one (*C. tamandua*), occur exclusively in water bodies within the Congo basin, with peaks of diversity observed in the rapids of the lower Congo region (Stewart and Roberts 1976). The most recent taxonomic revision of the genus (Poll et al. 1982) acknowledged 14 species, however this is likely to be an underestimate of the actual *Campylomormyrus* diversity, since many areas of the Congo basin have not been explored yet.

Recent molecular studies (Feulner et al. 2006, 2007) revealed that several described species, which are distinct in terms of morphology and EOD shape and length, constitute monophyletic groups. However, the inferred trees showed lack of resolution for two morphologically distinct taxa (*C. compressirostris, C.curvirostris*) and conflicting topologies between mitochondrial (*cytb*) and nuclear (*rps7*) gene trees. In the present study we analyzed 14 sympatric members of *Campylomormyrus*; nine of the selected groups represent already described species, while the remaining five correspond to possibly cryptic species characterized by peculiar EODs (Fig. 1). We analyzed sequence data collected from 5 molecular markers (one mitochondrial and four nuclear genes) in order to build a species tree of the genus and to delimit species using coalescent-based methods. We additionally used 16 microsatellite markers as an independent line of evidence for the detection of reproductively isolated groups within *Campylomormyrus*. Finally, we applied Approximate Bayesian Computation (ABC) and Posterior Predictive Simulations (PPS) on both datasets to select the diversification model that fitted the data best and to check for allelic introgression among species.

Figure 1



MATERIALS AND METHODS

Taxon Sampling

Specimens were collected throughout three different expeditions in the Republic of the Congo (2004, 2006 and 2012); all the analyzed samples come from the Congo River rapids south of Brazzaville (see Supplementary Data for the geographic coordinates). After the catch, the EOD of each specimen was measured, under the same conductivity conditions, using a Tektronix TDS 3012B digital phosphor oscilloscope, preamplified via an ADA400A differential preamplifier (gain= 10 Volts; upper bandwidth= 3kHz). Other individuals were obtained from aquarium traders located in Kinshasa (Democratic Republic of the Congo) as living specimens. From all fish, fin clips for genetic analysis and standardized pictures for morphometric measurements were taken. A unique identifier was assigned to each caught sample (Supplemental Table 1).

Molecular Data

Genomic DNA was isolated from fin clips or liver tissue stored in ethanol or Queens Buffer using the Dneasy Blood & Tissue Kit (Qiagen). In total, five markers were PCR amplified: 1 mitochondrial (cytochrome b, *cytb*) and 4 nuclear (the ribosomal protein s7, *rps7* and three introns of the *scn4aa* gene). The amplified products were sequenced using an ABI PRISM 3130xl sequencer (Applied Biosystems), sample sizes for each gene are reported in Table 1. Primer sequences and PCR/Sequencing conditions are available in the Supplementary Section. The obtained sequences were edited with BioEdit v7.2.4 (Ibis Biosciences, Carlsbad, CA) and aligned using the MAFFT v7.017 (Katoh et al. 2002) plug-in implemented in Geneious Pro v7.1.0 (Biomatters Limited, Auckland, New Zealand). The most likely nucleotide substitution models for each marker were calculated using ModelGenerator v0.85 (Keane et al. 2006) and selected after the Bayesian Information Criterion (BIC; Schwarz 1978) (Table 1). In addition, 242 individuals were genotyped for a set of 16 microsatellite markers (Feulner et al. 2005); allele sizes were estimated using GeneMapper v4.0 (Applied Biosystems).

Species Tree Estimation

A species tree for the genus *Campylomormyrus* was built using the *BEAST method (Heled and Drummond 2010) implemented in BEAST v1.7.5 (Drummond et al. 2012). In total, 14 groups were defined, corresponding to the EOD types illustrated in Figure 1, the related species *Gnathonemus petersii* was used as outgroup. Of the 14 pre-defined EOD types: 6 correspond to fully described species that are morphologically distinguishable (*C. numenius*, *C. rhynchophorus*, *C. curvirostris*, *C.*

christyi, *C. tamandua*); 1 is formed by a group whose morphology is not attributable to any available holotype (*C. sp. nov*, see also Feulner et al. 2007) and 7 are constituted by sets of individuals displaying different EODs within fully described species (*C. compressirostris*-triphasic, *C. compressirostris*-biphasic; *C. elephas, C. cf. elephas*; *C. alces*-short, *C. alces*-long, *C. alces*-triphasic). The main parameters used in the *BEAST analysis are reported in Table 1. The MCMC chain was run for 2 x 10^7 generations, parameters were sampled from the posterior distribution every 5,000 steps. After the run, chain stationarity was assessed using Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). TreeAnnotator v1.7.5 was used to obtain the maximum clade credibility tree from the collection of sampled species trees.

The *BEAST method implements the multispecies coalescent model and considers coalescent stochasticity as the only source of gene tree discordance. However, many other factors may result in discrepancies among gene trees (e.g., hybridization, selection, gene duplication/loss and recombination); it is therefore strongly recommended to test the fit of the obtained data to the underlying model. We used the R package starbeastPPS (Reid et al. 2013) to assess the fit of our data to the multispecies coalescent model. StarbeastPPS applies Posterior Predictive Simulations (PPS, Gelman et al. 2009) in order to produce a predictive distribution of coalescent trees simulated from the posterior distribution of species trees under the multispecies coalescent model. It uses then several test statistics to compare the empirical data with the posterior predictive distributions.

Species Delimitation

Several methods were applied on sequence and microsatellite data, separately, in order to detect species boundaries. A first analysis on sequence data was conducted using the program BPP v2.2 (Rannala and Yang 2003; Yang and Rannala 2010). This method accommodates the species phylogeny as well as lineage sorting due to ancestral polymorphism. The parameters in the model include the species divergence times τ, measured by the expected number of mutations per site, and population size parameters

$\theta=4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per site per generation, so that θ is the average proportion of different sites between two sequences sampled at random from the population. The program takes as input a guide-tree and a set of prior distributions for the parameters to be estimated. It then runs reversible-jump Markov Chain Monte Carlo simulations (rjMCMC, Rannala and Yang 2013) to explore the different parameter spaces across the alternative models of species delimitation and calculates posterior probability distributions (PP) of differentiation for the predefined groups on the guide-tree. The maximum clade

credibility species-tree obtained from *BEAST was chosen as the input guide-tree for this analysis. We applied the two algorithms available in the program 0 (ε=5 and 10) and 1 (α=2, m=1 and α=1, m=2) as recommended by the authors. Following previous studies (Leaché and Fujita 2010; Camargo et al. 2012; Ruane et al. 2013), we ran the analyses several times, using different prior values for θ and τ_0 (root age), in order to account for the possible combinations of population size and time divergence, by modifying the parameters (α, β) of the Γ prior distributions: large population size/deep divergence (both priors with α=2and β=2000), small population size/shallow divergence (both priors with α = 1 and β= 10), and large population size/shallow divergence (α = 2 and β = 2000 for θ prior; α = 1and β = 10 for τ prior). Each analysis was run twice, using different seed values, for 10 x 105 generations (thinning=2) with a burn-in of 2 x 104.

Besides the BPP method, we applied the discovery approach developed by Satler et al. (2013) and implemented in the software spedeSTEM (Ence and Carstens 2011) v2.0. This method utilizes STEM v2.0 (Kubatko et al. 2009) in order to infer the maximum likelihood species tree from a set of given gene trees, under all possible models of lineage composition; it then evaluates the log-likelihoods of the obtained species trees using information theory as described in Carstens and Dewey (2010) and selects the most likely model. The models of lineage composition compared in this study spanned from 14 (all EOD types correspond to independent lineages) to 1 (all samples belong to the same species). SpedeSTEM requires a set of pre-computed ultrametric gene trees and an overall estimate of θ as input data; the five gene trees were computed using BEAST v1.7.5, θ was calculated using the Watterson estimator (Watterson 1975) in Arlequin v3.5 (Excoffier and Lischer 2010).

Finally, we employed a single-locus species delimitation method: the General Mixed Yule Coalescent (GMYC, Pons et al. 2006) model, on our *cytb* data. The GMYC model assumes that the amount of lineage accumulation within species (coalescence process) occurs much faster than between species (Yule process) and aims at finding the transition point between the two processes on gene trees, in order to identify putative speciation events. Here, we used a bayesian version of the method, implemented in the R package bGMYC (Reid and Carstens 2012), which accounts for tree topology and branch lengths uncertainty by sampling from a posterior distribution of gene trees obtained using BEAST v1.7.5 (chain length = 1 x 108, sampled every 5 x 103 steps). We discarded 40% of the trees from the posterior distribution as burn-in and down-sampled the rest up to 100 trees; the analysis was run for 5 x 104 generations (burn-in=40,000; thinning=10).

In addition to sequence markers, we used genotypic data obtained from 16 microsatellite loci as an independent source of information to infer a partitioning scheme among the pre-defined groups. We first calculated pairwise F_{ST} values for the 14 EOD types by performing 1 x 10^4 permutations and setting the significance level to 0.05 and then executed an Analysis of Molecular Variance (AMOVA,

Excoffier et al. 1992) in Arlequin v3.5. For the AMOVA analysis we partitioned our dataset hierarchically by lumping the 14 EOD types into 10 groups, corresponding to morphologically distinct clusters. AMOVA null distributions were produced after 1.6 x 104 permutations ($p \leq 0.05$).

We also used STRUCTURE v2.3 (Pritchard et al. 2000) in order to identify putative reproductively isolated groups within the microsatellite dataset. The STRUCTURE algorithm works within a Bayesian framework and computes the most probable clustering scheme in the dataset given a fixed number of clusters (k). The analysis was run iteratively several times, applying different k values (number of assumed populations) for each iteration (k from 1 to 18) using the admixture model. The burn-in value for the MCMC chain was set to 2.5 x 104, the post burn-in chain was run for 1.5 x 105 generations; each single run was replicated 10 times with different seeds. The single optimal k value was estimated using the Evanno method (Evanno et al. 2005), implemented in the Structure Harvester tool (Earl and vonHoldt 2011). The replicated runs were summarized using CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) and the final data matrix was visualized with DISTRUCT v1.1 (Rosenberg 2004).

Model Selection

One of the main limitations of the species-tree inference (*BEAST) and species delimitation (BPP, spedeSTEM) methods that were applied on our sequence dataset is their utter reliance on the multispecies coalescent model. As previously mentioned, such model assumes no gene flow among populations and considers Incomplete Lineage Sorting (ILS) as the only process responsible for inconsistency among gene trees. However, migration among species cannot be excluded *a priori*, particularly when studying sympatric populations. Moreover, recent studies (Leaché et al. 2014) have demonstrated that unnoticed levels of migration may result in overestimates of population sizes and underestimates of divergence times on species-trees.

Here, we used Approximate Bayesian Computation (ABC, Beaumont et al. 2002) in order to test our dataset against two competing diversification models: i) a pure-isolation model (no migration among species) and ii) an isolation-with-migration model (IM, Nielsen and Wakeley 2001; Hey 2010), which contemplates a certain level of migration among populations during the speciation process. ABC techniques aim at approximating posterior distributions of parameters by: i) sampling the parameters of interest from the prior distributions; ii) simulating the data given the sampled parameters; iii) summarize the simulated data using a set of summary statistics (SuSt); iv) approximate the posterior distribution by retaining a fraction of the simulated data that falls within a pre-specified Euclidean distance from the real data (see Beaumont 2010). By avoiding the computation of complex likelihood

functions, ABC methodologies allow to infer parameter values from rather complicated models at the cost of some approximation. Here, we used popABC v1.0 (Lopes et al. 2009) to simulate data from the parameter priors and calculate a set of 13 summary statistics (see Supplementary material) under the pure-isolation and the isolation-with-migration models (1 x 105 simulation runs). The posterior probabilities of the two alternative models were then evaluated with the R package "abc" v1.5 (Csilléry et al. 2012) using the postpr function and the multinomial logistic regression method ($p \leq 0.05$). For this analyses a subtree from the whole species-tree was selected, lumping together those taxa that did not result to be reproductively isolated using the above mentioned species delimitation methods; both sequence and microsatellite data were used together to calculate the summary statistics.

Hybridization

Allelic introgression is another important factor which is known to influence gene tree topologies (Besansky and Powell 1994; Morando et al. 2004). Discerning introgression from incomplete lineage sorting (ILS) allows the identification of patterns of interspecific hybridization. In order to ascertain the possible role of mitochondrial introgression within

Campylomormyrus, we used the software JML v1.02 (Joly 2012). This program uses posterior predictive checking to test whether the minimum pairwise distances between the sequences of two species are smaller than expected under a scenario that does not account for hybridization. It then applies a one-tailed test to check if the observed distances are significantly smaller than the set of distances obtained from simulated data. JML uses as input posterior distributions of species trees, population sizes and divergence times. Species trees were sampled from a posterior distribution obtained by running *BEAST on a subset of 6 species *(C. compressirostris, C.curvirostris, C. rhynchophorus, C. numenius, C. tshokwe, C. sp. nov.)*; the MCMC chain length was set to 2×10^8, parameters were sampled every 5×10^3 steps. *Cytb* datasets were simulated from the species-tree distributions (burn-in=10%), the significance value was set to 0.05.

In order to confirm these patterns of introgression, we looked for admixture signatures in our microsatellite dataset. We employed the software DIYABC v2.0 (Cornuet et al. 2014) to perform model selection among three historical diversification scenarios involving the same above mentioned 6 species: i) scenario 1, no admixture; ii) scenario 2, admixture between *C. compressirostris* and *C. rhynchophorus* iii) scenario 3, admixture between *C. rhynchophorus* and *C.curvirostris*. DIYABC makes use of ABC to estimate the posterior probabilities of parameters by simulating datasets and computing a set of summary statistics. Unlike popABC, DIYABC simulates data under a Wright-

Fisher model, meaning that no migration is involved. For each scenario 3 x 10^6 simulations were run and 12 summary statistics calculated (see Supplementary material).

RESULTS

Molecular Data

Sequence information concerning the five molecular markers used in this study is summarized in Table 1. All the sequences are accessible from the GenBank database under the Accession Numbers: KJ680159-KJ680224 (*scn4a*, intron 15); KJ713702-KJ713777 (*scn4a*, intron 14); KJ713778-KJ713851 (*scn4a*, intron 1); KJ713852-KJ713893 (*rps7*); KJ713894-KJ713964 (*cytb*). They increase the existing amount of molecular information for the *Campylomormyrus* genus by both expanding the current *cytb* and *rps7* taxonomic coverage to three additional described species (*C. christyi, C.elephas* and *C.alces*) and by adding three new nuclear markers to the present collection of molecular data.

With regard to microsatellite data, the information gathered in this study also augments the present taxonomic range (*C. christyi, C.elephas, C.alces* and *C. sp. nov.*) compared to previous ones (Feulner et al. 2006, 2007). Allele sizes (including the flanking region) for each genotyped individual are reported in Supplemental Table 2.

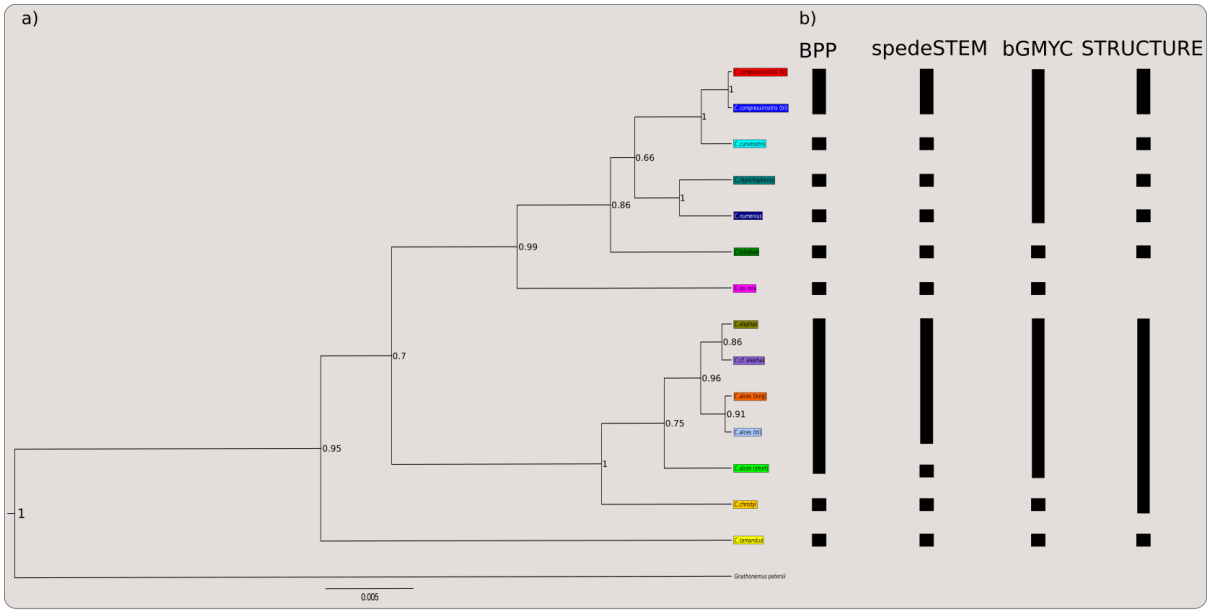
Species Tree Inference

The ultrametric *BEAST species tree is illustrated in Figure 2a, it indicates *C. tamandua* as the most basal lineage within *Campylomormyrus* and displays then two distinct macroclades: one formed by all the members of the so-called "*alces*" complex of species (*C. alces*, *C.christyi* and *C.elephas*) and the other constituted by all the remaining *Campylomormyrus* members. Within the "*alces*" clade *C. christyi* forms the most basal lineage. Particularly interesting here is the position of *C. alces* (short) which does not form a monophyletic group with *C. alces* (long) and *C. alces* (tri). Inside the second macroclade the two species characterized by a very long EOD (~ 40ms), *C. numenius* and *C. rhynchophorus*, form a monophyletic group.

The results of the PPS analysis show the fit of our empirical data to the multispecies coalescent model. Figure 3 displays the position of the observed gene trees relative to the null distributions obtained from the simulated ones. The fit to the multispecies coalescent holds both in terms of

coalescent likelihoods and expected number of deep coalescences for all loci, considered either together or separately.

Figure 2

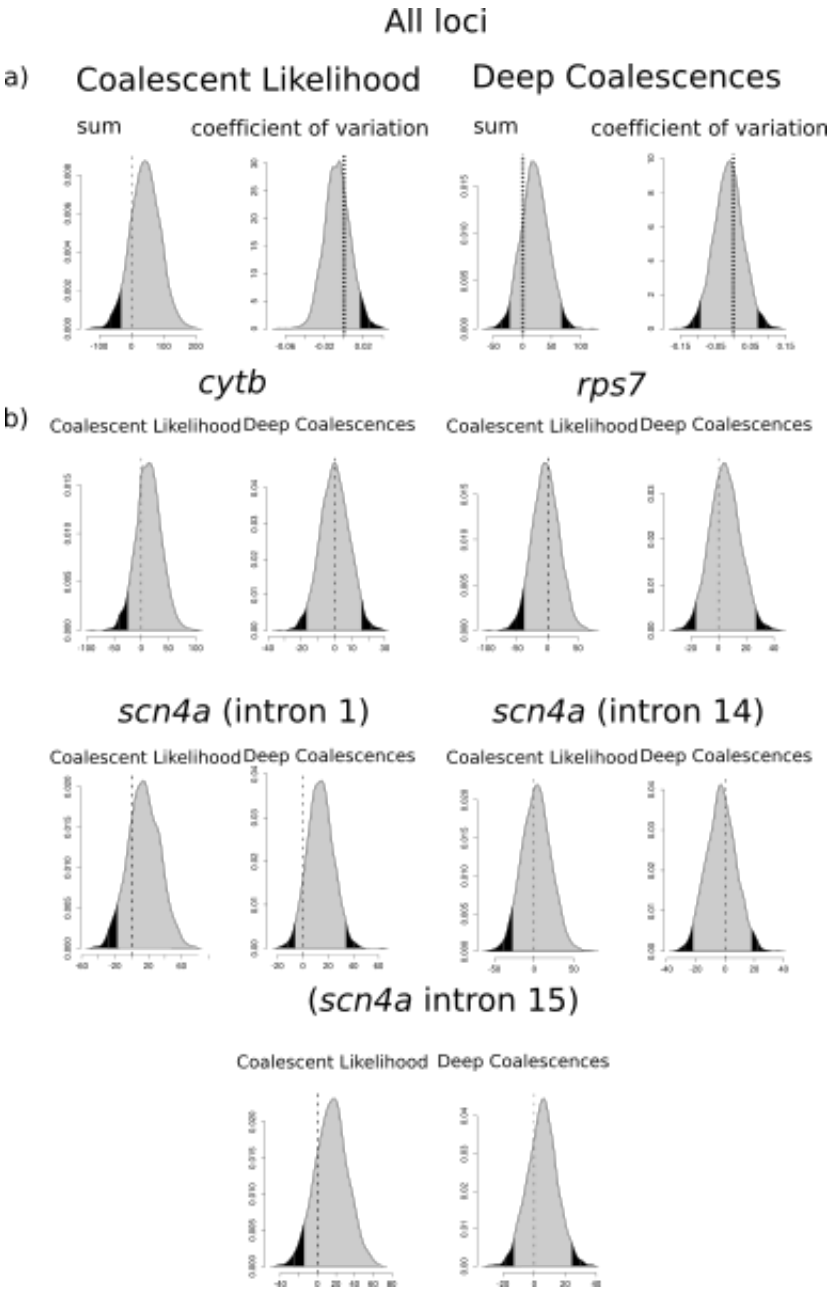


Species Delimitation

The results obtained from the different applied species delimitation methods are summarized in Figure 2b. The BPP analysis run on the *BEAST guide tree identified 9 lineages out of 14 as distinct species using stringent posterior probability values (PP ≥ 0.95). The same pattern holds over the different combinations of population size and tree depth priors used in the analysis (see Supplementary material). The two *C. compressirostris* EOD types and all the members of the "*alces*" complex, except C. christyi, displayed low posterior probability values. Using the spedeSTEM method, the species tree composed by 10 lineages had the highest support value (wi = 0.44, Table 2). These results retrace those obtained from BPP, with the only difference that *C. alces* (short) constitutes an additional separate species within the "*alces*" group. The results achieved using single locus information (*cytb*) with bGMYC detected an average of 6 independent lineages (coalescent units), with 95% Highest Posterior Density HPD distribution spanning from 2 to 33 units.

Additional tests of differentiation among groups were executed on the microsatellite dataset. Pairwise FST values are reported in Supplemental Table 3. AMOVA results (Table 3) show an extensive

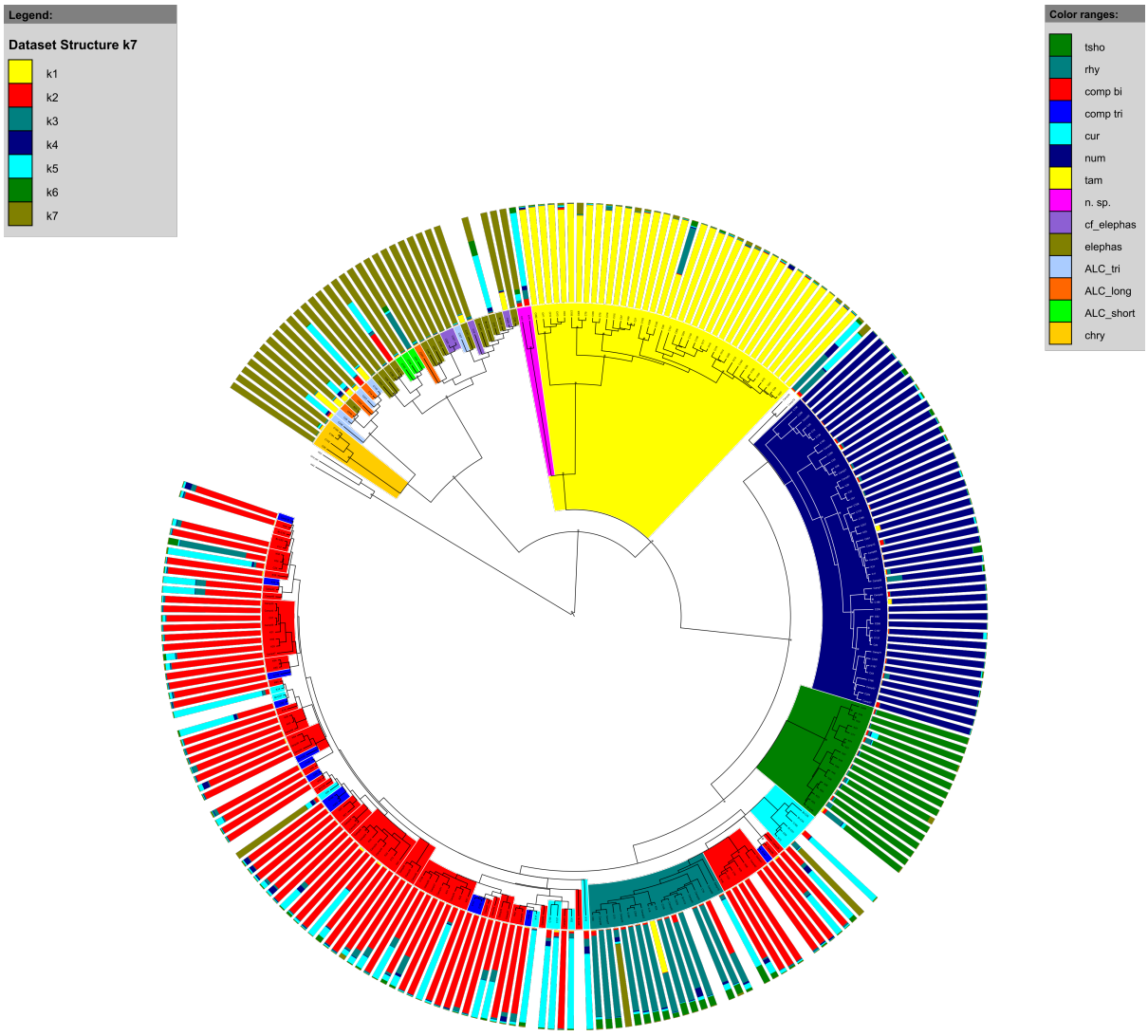
Figure 3



amount of population structure occurring mainly at the "among-species" hierarchical level. Furthermore, a small, but significant, level of differentiation is observed among EOD types within species. The outcome of the structure analysis suggests two possible clustering patterns (k=9; k=7), which displayed similar Δk values after the application of the Evanno method (Supplemental Figure 1). The two partitioning schemes are almost identical, with the two additional clusters in k=9 suggesting a certain level of population substructure within *C. compressirostris*. In both cases the identified clusters correspond to the already described species, with the only exception constituted by

all the members of the "*alces*" group, which are all grouped together. None of the within-species EOD types clustered separately.

Figure 4

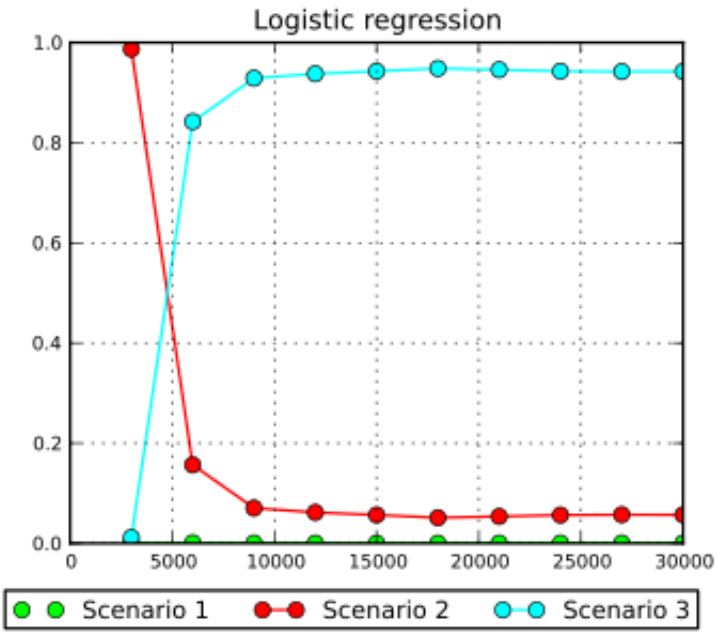


Testing models of diversification and hybridization.

We applied ABC-based model selection in order to test which of two competing diversification modalities (pure isolation; isolation with migration) fit the data best. The whole analysis was performed on a subtree of the *BEAST species-tree, consisting of 6 taxa (*C. sp. nov, C. tshokwe, C. rhynchophorus, C. numenius, C. curvirostris, C. compressirostris*). The results of the simulations largely support the isolation-with-migration scenario (PP=0.99), suggesting that a certain level of gene flow occurred among the diverging populations during the speciation process.

The same set of taxa was tested for possible signatures of mitochondrial introgression using the *cytb* dataset. The results of the posterior predictive checks, acquired using JML, revealed that the observed pairwise distances between the couples *C. rhynchophorus*-*C. compressirostris* and *C. rhynchophorus*-*C. curvirostris* were significantly smaller than those expected under a no admixture model (Supplemental Table 4). Mitochondrial introgression is also suggested by the position of *C. rhynchophorus* in the *cytb* gene-tree (Fig. 4), where it forms a monophyletic group placed within the unresolved *C. compressorostris*-*C. curvirostris* clade. The pattern of admixture observed for the mitochondrial dataset is confirmed by microsatellite data. In fact, the results of the DIYABC analysis clearly support the two admixture scenarios (*C. rhynchophorus*-*C. compressirostris* and *C. rhynchophorus*-*C. curvirostris*) over the no admixture one (Fig. 5).

Figure 5



DISCUSSION

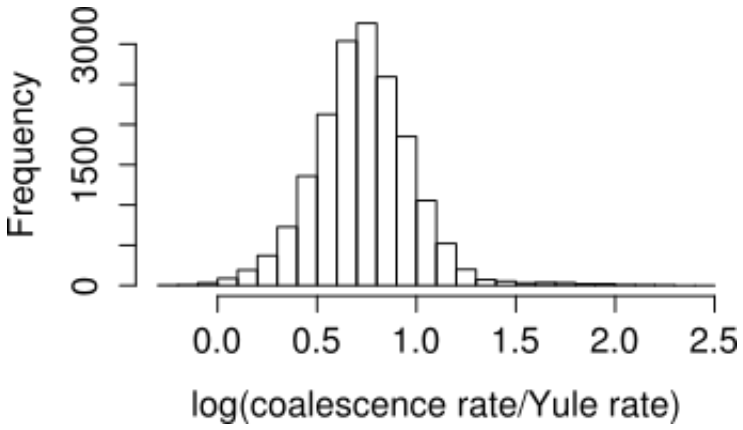
Species Tree and Species Delimitation

The combined application of coalescent based species-tree estimation with different species delimitation methodologies to the genus *Campylomormyrus* revealed the presence of two main species assemblages; the first one formed by: *C. compressirostris* (biphasic), *C. compressirostris* (triphasic), *C. curvirostris*, *C. rhynchophorus*, *C. numenius*, *C. tshokwe*, *C. sp. nov.*; the second one represented by all the members of the "*alces*" complex. In the first group all the taxa that can be morphologically discriminated resulted to be reproductively isolated (see Figure 2a); conversely, the two EOD types within *C. compressirostris* did not show any sign of separation by any of the used methods. An interesting achievement of the methods used in this study is the separation of *C. curvirostris* from *C. compressirostris*; this two species, in fact, result to be completely intertwined at the gene tree level (Fig. 4 and Supplemental Figures 3-6). Additionally, our results suggest that *C. sp. nov.* can be effectively considered as a distinct, yet undescribed species. Different is the situation regarding the "*alces*"-complex. Despite the fact that the three species which form the complex (*C. alces*, *C. elephas* and *C. christyi*) can be distinguished in terms of few anatomical characters (e.g. snout shape) and that within *C. alces* and *C.elephas* different EOD types are present, the whole group does not show any sign of differentiation when the microsatellites are considered, whereas only *C. christyi* constitutes an independent lineage using the data from the five DNA sequence markers. These results might be indicative of very rapid speciation events, possibly driven by strong selection acting on loci involved in morphological diversification. Indeed, simulation experiments performed on BPP, found that this method cannot distinguish among separate lineages that diverged less than $0.4N_e$ generations ago using 5 loci and $\theta = 10$ (Yang and Rannala 2010). Increasing the sampling effort to more loci and more sequences per population would help to understand more about species boundaries within this complex. The presence of divergent EOD types within sympatric populations, characterized by striking levels of genetic similarity has been already observed in other mormyrid species (Arnegard et al. 2005). Such pattern of diversification could be explained as the effect of incipient sympatric speciation events. If we assume that the phenotype "EOD" is subject to a certain level of, direct or indirect, natural selection (Coyne and Orr 2004), acting on one or few loci, the markers used in this study may not be sufficient to capture any signature of reproductive isolation due to the exclusive or combined effects of recombination and large effective population sizes (Noor and Feder 2006; Nosil and Schluter 2011; Butlin et al. 2012). An alternative scenario would invoke the role of phenotypic plasticity in determining EOD variability. Since EOD features are mainly determined by the anatomy of the electric organ (Westby and Kirschbaum 1978; Denizot et al. 1982), environmental factors acting on a certain ontogenetic stage might influence EOD diversification (Jia et al. 2000; Rodriguez and Greenfield 2003; Miner et al. 2005).

The application of the bGMYC delimitation method on the *cytb* dataset retrieved the least number of independent lineages (5), such low performance is almost certainly due to the high amount of uncertainty among gene trees, particularly at deeper nodes (Supplemental Figure 2), that in turn

reflects the occurrence of substantial levels of incomplete lineage sorting and mitochondrial introgression (see below) and results in a low ratio between coalescence and Yule lineage accumulation rates (Fig. 6).

Figure 6



Species-tree inference and species delimitation methods based on the multispecies coalescent model constitute very valid tools of investigation. Nevertheless, violations of their underlying assumptions may introduce inference errors. Simulation studies have demonstrated the robustness of the model against certain kinds of violations (e.g. gene flow among species; Ence and Carstens 2011; Zhang et al. 2011; Heled et al. 2013). However, when working with empirical data the consistency between the data and the model is not

aprioristically known; for this reason the application of methods that can assess such fit is of vital importance. The usage of PPS model-checking techniques on the *BEAST species-tree reveals the coherence between the 5 markers developed for this study and the model, this does not necessarily mean that the topological conflicts among gene trees are the pure result of ILS (one of the main assumptions of the multispecies coalescent model), but that, if any other biological process (e.g. natural selection, hybridization, gene duplication) has shaped the present pattern of observed data, it should not drastically influence the outcome of an analysis based on such model.

Diversification Modalities

The application of ABC-based model selection on a subset of 6 *Campylomormyrus* species supports isolation-with-migration over pure isolation as the most probable model of diversification. This piece of evidence must be taken into account when interpreting parameters like speciation times and population sizes, because the occurrence of gene flow among independent lineages is known to introduce biases in the estimation of these parameters (Leaché et al. 2013). Furthermore, it is

compatible with the hypothesis of a sympatric origin for this group of species (Feulner et al. 2006, 2007), although additional evidence must be obtained by: i) exploring the extent of geographic variation among allopatric populations; ii) investigating genome-wide patterns of disruptive selection and their level of linkage to loci responsible for assortative mating (e.g. EOD) and testing whether they can overcome the homogenizing effects of gene flow (Felsenstein 1981; Bolnick and Fitzpatrick 2007; Pinho and Hey 2010).

Hybridization

The analyses conducted on the *cytb* dataset indicate the presence of mitochondrial introgression between *C. rhynchophorus* and members of the clade composed by *C. curvirostris* and *C. compressirostris*. This line of evidence is corroborated by microsatellite analyses showing: i) highest posterior probability for the scenario involving admixture between *C. rhynchophorus* and *C. curvirostris* (Fig. 5); ii) presence of possibly "admixed" genotypes in the STRUCTURE output (Fig. 4).

These findings suggest a certain level of "permeability" to gene-flow in characters involved in pre-zygotic isolation like EODs. Choice experiments using female *C. compressirostris* and male *C. rhynchophorus* specimens revealed clear preference for conspecifics, either when actual specimens or only recorded EODs were used, indicating the effectiveness of species-specific electric signals in maintaining assortative mating (Feulner et al. 2009b). A possible explanation for such seemingly conflicting results may be found by invoking the role of frequency-dependent mate-choice – i.e., the occurrence of heterospecific matings when one of the two species cannot find any conspecific - determining the observed levels of introgression (Chan and Levin 2005). Setting no-choice behavioral experiments using different species would shed light on the actual possibility of this mechanism in nature (Seehausen 1997; Knight et al. 1998).

CONCLUSIONS

The results achieved in this study indicate the importance of utilizing methods for phylogenetic inference and species delimitation that explicitly model the sources of discordance and uncertainty among gene trees. The use of coalescent-based species delimitation methods partially succeeded in unraveling the species boundaries within the genus *Campylomormyrus*. However, species groups that

appear to display more complex evolutionary histories ("*alces*"-complex) deserve more attention in the future by increasing the sampling effort both in terms of individuals and loci.

Finally, the identification of the isolation-with-migration as a putative diversification scenario constitutes a valid starting hypothesis that needs to be further evaluated, in order to better understand the mechanisms of speciation in weakly-electric fish.

SUPPLEMENTARY MATERIAL

Supplementary figures, tables and data can be found at http://datadryad.org in the Dryad repository (doi:).

FUNDING

This study was funded by the University of Potsdam and by the Leibniz-SAW-project GENART.

REFERENCES

- Alves-Gomes J., Hopkins C.D. 1997. Molecular Insights into the Phylogeny of Mormyriform Fishes and the Evolution of Their Electric Organs; pp. 343–351. Brain. Behav. Evol. 49:343–351.
- Arnegard M.E., Bogdanowicz S.M., Hopkins C.D. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. Evolution. 59:324– 43.
- Barluenga M., Stölting K.N., Salzburger W., Muschick M., Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature. 439:719–23.
- Bastian J. 1994. Electrosensory Organisms. Phys. Today. 47:30-37.
- Beaumont M. A. 2010. Approximate Bayesian Computation in Evolution and Ecology. Annu. Rev. Ecol. Evol. Syst. 41:379–406.

- Beaumont M.A., Zhang W., Balding D.J. 2002. Approximate Bayesian Computation in Population Genetics. Genetics. 162:2025–2035.
- Besansky N., Powell J. 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proc. Natl. Acad. Sci. U. S. A. 91:6885–6888.
- Bolnick D.I., Fitzpatrick B.M. 2007. Speciation : Sympatric Models and Empirical Evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.
- Bratton B.O., Kramer B. 1989. Patterns of the electric organ discharge during courtship and spawning in the mormyrid fish, Pollimyrus isidori. Behav. Ecol. Sociobiol. 24:349–368.
- Bullock T., Fay R., Hopkins C., Popper A. 2005. Electroreception. New York: Springer.
- Butlin R., Debelle A., Kerth C., Snook R.R., Beukeboom L.W., Castillo Cajas R.F., Diao W., Maan M.E., Paolucci S., Weissing F.J., van de Zande L., Hoikkala A., Geuverink E., Jennings J., Kankare M., Knott K.E., Tyukmaeva V.I., Zoumadakis C., Ritchie M.G., Barker D., Immonen E., Kirkpatrick M., Noor M., Macias Garcia C., Schmitt T., Schilthuizen M. 2012. What do we need to know about speciation? Trends Ecol. Evol. 27:27–39.
- Camargo A., Morando M., Avila L.J., Sites J.W. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the Liolaemus darwinii complex (Squamata: Liolaemidae). Evolution. 66:2834–49.
- Camargo A., Sites J.J. 2013. Species Delimitation: A Decade After the Renaissance. In: Pavlinov I.Y., editor. The species problem - ongoing issues. InTech. p. 225–247.
- Carstens B.C., Dewey T. A. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. Syst. Biol. 59:400–14.
- Chan K.M.A., Levin S.A. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution. 59:720–9.
- Cornuet J., Pudlo P., Veyssier J., Dehne- Garcia A., Gautier M., Leblois R., Marin J., Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics. 30:1187–1189.
- Coyne J.A., Orr H.A. 2004. Speciation. Sunderland, Massachusets U.S.A.: Sinauer Associates, Inc.
- Crawford J.D. 1991. Sex Recognition by Electric Cues in a Sound-Producing Mormyrid Fish, *Pollimyrus isidori*. Brain. Behav. Evol. 38:20–38.

- Csilléry K., François O., Blum M.G.B. 2012. abc: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3:475–479.
- Daget J., Gosse J.P., Teugels G.G., Thys van den Audenaerde D.F.E. 1991. Check list of the freshwater fishes of Africa. Paris, Bruxelles: ORSTOM.
- Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.
- Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–40.
- Denizot J.P., Kirschbaum F., Westby G.W.M., Tsuji S. 1982. On the development of the adult electric organ in the mormyrid fish Pollimyrus isidori (with special focus on the innervation). J. Neurocytol. 11:913–934.
- Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–73.
- Earl D.A., vonHoldt B.M. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.
- Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. U. S. A. 104:5936–41.
- Von der Emde G. 1999. Active electrolocation of objects in weakly electric fish. J. Exp. Biol. 202:1205–1215.
- Ence D.D., Carstens B.C. 2011. SpedeSTEM: a rapid and accurate method for species delimitation. Mol. Ecol. Resour. 11:473–80.
- Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–20.
- Excoffier L., Lischer H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–7.
- Excoffier L., Smouse P.E., Quattro J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131:479–91.
- Felsenstein J. 1981. Skepticism Towards Santa Rosalia , or Why are There so Few Kinds of Animals ? Evolution. 35:124–138.
- Feulner P., Kirschbaum F., Tiedemann R. 2008. Adaptive radiation in the Congo River: An ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; Campylomormyrus). J. Physiology-Paris. 102:340–346.

- Feulner P.G.D., Kirschbaum F., Mamonekene V., Ketmaier V., Tiedemann R. 2007. Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: Campylomormyrus): a combined molecular and morphological approach. J. Evol. Biol. 20:403–414.
- Feulner P.G.D., Kirschbaum F., Schugardt C., Ketmaier V., Tiedemann R. 2006. Electrophysiological and molecular genetic evidence for sympatrically occuring cryptic species in African weakly electric fishes (Teleostei: Mormyridae): Mol. Phylogenet. Evol. 39:198–208.
- Feulner P.G.D., Kirschbaum F., Tiedemann R. 2005. Eighteen microsatellite loci for endemic African weakly electric fish (Campylomormyrus, Mormyridae) and their cross species applicability among related taxa. Mol. Ecol. Notes. 5:446–448.
- Feulner P.G.D., Plath M., Engelmann J., Kirschbaum F., Tiedemann R. 2009a. Magic trait Electric Organ Discharge (EOD): Dual function of electric signals promotes speciation in African weakly electric fish. Commun. Integr. Biol. 2:329–331.
- Feulner P.G.D., Plath M., Engelmann J., Kirschbaum F., Tiedemann R. 2009b. Electrifying love: electric fish use species-specific discharge for mate recognition. Biol. Lett. 5:225– 228.
- Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–8.
- Gelman A., Carlin J.B., Stern H.S., Rubin D.B. 2009. Bayesian Data Analysis. Chapman & Hall/CRC.
- Hausdorf B. 2011. Progress toward a general species concept. Evolution. 65:923–31.
- Heled J., Bryant D., Drummond A.J. 2013. Simulating gene trees under the multispecies coalescent and time-dependent migration. BMC Evol. Biol. 13:44.
- Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–80.
- Hey J. 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27:905–20.
- Higgie M. 2000. Natural Selection and the Reinforcement of Mate Recognition. Science. 290:519–521.
- Jakobsson M., Rosenberg N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23:1801–6.
- Jia F., Greenfield M.D., Collins R.D. 2000. Genetic Variance of Sexually Selected Traits in Waxmoths: Maintenance by Genotype X Environment Interaction. Evolution. 54:953– 967.
- Jiggins C.D., Naisbit R.E., Coe R.L., Mallet J. 2001. Reproductive isolation caused by colour pattern mimicry. Nature. 411:302–5.
- Joly S. 2012. JML: testing hybridization from species trees. Mol. Ecol. Resour. 12:179–84.

- Katoh K., Kazuharu M., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
- Keane T.M., Creevey C.J., Pentony M.M., Naughton T.J., Mclnerney J.O. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. 6:29.
- Knight M.E., Turner G.F., Rico C., Van Oppen M.J.H., Hewitt G.M. 1998. Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. Mol. Ecol. 7:1605–1610.
- Kramer B., Kuhn B. 1994. Species recognition by the sequence of discharge intervals in weakly electric fishes of the genus Campylomormyrus (Mormyridae, Teleostei). Anim. Behav. 48:435–445.
- Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics. 25:971–3.
- Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.
- Lamml M., Kramer B. 2006. Differentiation of courtship songs in parapatric sibling species of dwarf stonebashers from southern Africa (Mormyridae , Teleostei). Behaviour. 143:783–810.
- Lavoué S., Sullivan J.P., Arnegard M.E., Hopkins C.D. 2008. Differentiation of morphology, genetics and electric signals in a region of sympatry between sister species of African electric fish (Mormyridae). J. Evol. Biol. 21:1030–45.
- Leaché A.D., Fujita M.K. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). P. Roy. Soc. Lond. B Bio. 277:3071–7.
- Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.
- Lissmann H.W., Machin K.E. 1958. The Mechanism of Object Location in *Gymnarchus niloticus* and Similar Fish. J. Exp. Biol. 35:451–486.
- Lissmann H.W. 1951. Continuous Electrical Signals from the Tail of a Fish, *Gymnarchus niloticus* Cuv. Nature.167:201–202.
- Lissmann H.W. 1958. On the Function and Evolution of Electric Organs in Fish. J. Exp. Biol. 35:156–191.
- Lopes J.S., Balding D., Beaumont M. a. 2009. PopABC: a program to infer historical demographic parameters. Bioinformatics. 25:2747–9.
- Marrero C., Winemiller K. 1993. Tube-snouted gymnotiform and mormyriform fishes: convergence of a specialized foraging mode in teleosts. Environ. Biol. Fishes.299–309.

- Miner B.G., Sultan S.E., Morgan S.G., Padilla D.K., Relyea R.A. 2005. Ecological consequences of phenotypic plasticity. Trends Ecol. Evol. 20:685–92.
- Morando M., Avila L.J., Baker J., Sites J.W. 2004. Phylogeny and Phylogeography of the *Liolaemus darwinii* Complex (Squamata: Liolaemiidae): Evidence for Introgression and Incomplete Lineage Sorting. Evolution. 58:842–859.
- Nielsen R., Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics. 158:885–96.
- Noor M., Feder J. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7:851-861.
- Nosil P., Schluter D. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160–167.
- O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for Different Rates of Continuous Trait Evolution Using Likelihood. Evolution. 60:922–933.
- Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–83.
- Pinho C., Hey J. 2010. Divergence with Gene Flow: Models and Data. Annu. Rev. Ecol. Evol. Syst. 41:215–230.
- Poll M., Gosse J., Orts S. 1982. Le genre Campylomormyrus Bleeker, 1874, étude systematique et description d'une espece nouvelle (Pisces, Mormyridae). Bull. l'Institut R. des Sci. Belgique Biol. 54:1–34.
- Pons J., Barraclough T., Gomez-Zurita J., Cardoso A., Duran D., Hazell S., Kamoun S., Sumlin W., Vogler A. 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Syst. Biol. 55:595–609.
- Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.
- De Queiroz K. 1998. The General Lineage Concept of Species, Species Criteria, and the Process of Speciation. Endless Forms: Species and Speciation. Oxford University Press. p. 57–75.
- De Queiroz K. 2011. Branches in the lines of descent: Charles Darwin and the evolution of the species concept. Biol. J. Linn. Soc. 103:19–35.
- Rannala B., Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics. 164:1645–1656.
- Rannala B., Yang Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics. 194:245–53.
- Reid N.M., Carstens B.C. 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evol. Biol. 12:196.

- Reid N.M., Hird S.M., Brown J.M., Pelletier T. a, McVay J.D., Satler J.D., Carstens B.C. 2013. Poor Fit to the Multispecies Coalescent is Widely Detectable in Empirical Data. Syst. Biol. 63:322–333.
- Rodríguez R.L., Greenfield M.D. 2003. Genetic Variance and Phenotypic Plasticity in a Component of Female Mate Choice in an Ultrasonic Moth. Evolution. 57:1304–1313.
- Rosenberg N. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes. 4:137-138
- Ruane S., Bryson R.W., Pyron R.A., Burbrink F.T. 2013. Coalescent Species Delimitation in Milksnakes (genus Lampropeltis) and Impacts on Phylogenetic Comparative Analyses. Syst. Biol. 63:231–250.
- Satler J.D., Carstens B.C., Hedin M. 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (mygalomorphae, antrodiaetidae, aliatypus). Syst. Biol. 62:805–23.
- Schwarz G. 1978. Estimating the Dimension of a Model. Ann. Stat. 6:461–464.
- Seehausen O. 1997. Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. Science. 277:1808–1811.
- Simpson G.G. 1951. The Species Concept. Evolution. 5:285–298.
- Stewart D.J., Roberts T.R. 1976. An Ecological and Systematic Survey of Fishes in the Rapids of the Lower Zaire or Congo River. Bull. Mus. Comp. Zool. 147:239–317.
- Tajima F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics.437–460.
- Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol. Evol. 16:381–390.
- Watterson G.A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7:256–276.
- Westby G.W.M., Kirschbaum F. 1978. Emergence and Development of the Electric Organ Discharge in the Mormyrid Fish , *Pollimyrus isidori*. II. Replacement of the Larval by the Adult Discharge. J. Comp. Physiol. A.45–59.
- Wilson I.J., Balding D.J. 1998. Genealogical inference from microsatellite data. Genetics. 150:499–510.
- Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U. S. A. 107:9264–9.Z
- Zhang C., Zhang D., Zhu T., Yang Z. 2011. Evaluation of a bayesian coalescent method of species delimitation. Syst. Biol. 60:747–61.

FIGURE 1. Specimens analyzed in this study and their EODs. The EODs are reported on the same time scale to emphasize the pronounced difference in duration in some species. The signals embedded in circles were magnified 10x in order to show their shape features.

FIGURE 2. (a) Species tree estimated using *BEAST, colored labels correspond to the color code used for EODs in Figure 1. (b) Results of the 4 applied specie delimitation methods. Black bars indicate species groupings on the species tree.

FIGURE 3. PPS check results. Distribution of test statistics for all loci considered together (a) and separately (b). Black dashed line correspond to the observed results; black shadings indicate the boundaries of the 95% highest posterior predictive density intervals.

FIGURE 4. *Cytb* gene tree (inner circle) and STRUCTURE plot showing the clustering results relative to the 16 microsatellite loci (outer circle).

FIGURE 5. Results of the DIYABC analysis. Scenario 1= No admixture; Scenario 2= admixture between *C. compressirostris* and *C. rhynchophorus*; Scenario 3= admixture between *C. rhynchophorus* and *C.curvirostris*. The x-axis reports the number of simulated datasets.

FIGURE 6. Distribution of ratios of the Coalescence to Yule rates sampled in the bGMYC analysis

8 Discussion

8.1 Transcriptomic analysis of electric organ and skeletal muscle

Electric organs have fascinated naturalists for thousands of years. Indeed, the earliest recorded illustrations of mormyrid fishes are depicted on ancient Egyptian tombstones and relics (Kellaway 1946). Even Darwin in "*On the Origin of Species*" (Darwin 1859) considered them as particularly challenging for his theory, in fact he found '*impossible to conceive by, what steps these wondrous organs have been produced*'. We now know that electric organs have evolved several times independently for the purpose of electro-location (Lissmann & Machin 1958) and electro-communication (Lissmann 1958). Since the pioneering work of Lissmann, the histology, anatomy and neurophysiology of electric organs have been thoroughly investigated. Despite the large amount of information gathered by these studies, no large-scale molecular datasets were produced yet. In this study I tried to fill this gap by: i) producing extensive transcriptomic resources for the electric organ and the skeletal muscle, and ii) identifying differentially expressed genes between these two tissues.

8.1.1 Sequencing and characterization of reference transcriptomes

African weakly electric fishes are non-model organisms; in fact, all teleost fishes with fully sequenced genomes (e.g. *Danio rerio*) belong to the taxon Clupeocephala, which separated from the Osteoglossomorpha approximately 300 Mya (Near *et al.* 2012). Additionally, none of the model species available possess an electric organ. For this reason a reference transcriptome is an essential prerequisite for the genetic investigation of an evolutionary novelty like the electric organ. Therefore, the first efforts of this study were devoted to the production of reference transcriptomes for the electric organ and skeletal muscle of *C. compressirostris*. These first sequencing experiments have set the first large-scale transcriptome datasets for a weakly-discharging electric fish. They have demonstrated that it is possible to sequence, assemble and annotate a transcriptome *de novo*, for explorative purposes, using only single-end, short-read (100bp) sequencing technologies.

The amount of matches against the Unigene database, 16 978 (electric organ) and 12 076 (skeletal muscle), over a total of 26 206 for *Danio rerio*, suggests a good coverage in terms of transcriptome diversity, despite the use of a non-normalized library. The different number of sampled genes might be explained by the fact that in muscle tissue there are several specific components – for example myosin and creatine kinase – that are massively expressed, in order to maintain the functional role of the organ, when using a non-normalized library, these highly expressed genes account for a huge number of reads, which can prevent the detection of rarer transcripts.

Given the explorative nature of this sequencing experiment, only qualitative comparisons could be done among the two transcriptomes. These comparisons were mainly based on the identification of

patterns of presence/absence of candidate genes between the two tissues. Several of these patterns were later confirmed by more rigorous, quantitative analyses of differentially expressed genes (see next paragraph). Among them it is worth mentioning various voltage-gated ion channels (e.g. *scn4aa*) and genes involved in neuronal development (e.g. *ncam*). On the other hand, the presence of typical skeletal muscle transcripts within the electric organ's transcriptome (e.g. *myl10a*) is not fully surprising. In fact, the electric organ, although being a non-contractile organ, displays yet many disorganized myofibrils within its electrocytes (Denizot *et al.* 1982).

Beside the acquisition of functional information, transcriptome-based data can also be used for the development of molecular markers useful for downstream "wet-lab" applications. For instance, in the present study 1671 unique Short Tandem Repeat (SSR or microsatellites) markers were identified, together with their flanking regions. The primers designed from these markers will contribute to massively augment the present set of available microsatellites (18; Feulner *et al.* 2005) for *Campylomormyrus*, allowing to conduct further population genomics investigations on this genus. Moreover, the annotation and retrieval of partial or complete Open Reading Frames (ORFs) on most candidate genes will be utilized for the development of primers to be employed in qPCR experiments that will be used for testing the patterns of gene expression obtained in this study.

The whole mitochondrial genome of *C. compressirostris* was also obtained; it represents the first fully-sequenced mito-genome for a member of the *Campylomormyrus* genus. It has been used in the course of this study for designing more specific and effective primers for the amplification of the cytochrome b (*cytb*) gene. In principle, it is not possible to obtain complete mitochondrial genomes from transcriptomic data. Indeed, there is a small portion of the control region that is not transcribed in vertebrates, because it is located between the two transcription promoters of the mito-genome (Shadel & Clayton 1997). The presence of this fragment in the reported mito-genome might have been caused by a small amount of genomic DNA contamination in the sequenced libraries.

8.1.2 Patterns of gene-expression between skeletal muscle and electric organ

As already stated, the electric organs of mormyrid fishes have evolved from skeletal muscle tissue involving the modification of portions of the caudal musculature. The main phenotypic differences that emerged in the electric organ from its muscular counterpart are: i) the loss of contractile activity; ii) the anatomical transformation of its basal unit from a tube-like structure –i.e., the myocyte– to a disc-like shape –i.e., the electrocyte; iii) the evolution of a complex and highly variable system of plasma membrane ex-vaginations and protrusions –i.e., the stalk system–, that influences considerably the electrical properties of the electrocyte (e.g., impedance), and iv) the presence of ion pumps and channels that are responsible for the production of the EOD. The combination of these characteristics determines the ability to produce an electric field by the electric organ.

In this part of my work, I wanted to investigate how such phenotypic differences are reflected in the respective transcriptomes of skeletal muscle and electric organ in *Campylomormyrus*. By exploring the overall patterns of gene expression between the two tissues in *C. compressirostris* and *C. tshokwe*, I have been able to identify which genes are differentially expressed and to functionally annotate them. Additionally, the detection of the commonalities in terms of gene expression patterns between the two species –i.e, the pool of genes that are differentially expressed in both species– is important for determining, which are the "core" genes that might be responsible for the differentiation of the electric organ from myogenic tissue.

From a technical point of view, this study has shown that it is possible to obtain high quality total RNA even from tissues sampled in the wild and under unfavorable conditions (e.g., tropical climate and absence of refrigerators), provided that they are handled appropriately (see the "Methods" section of Article II). The use of paired-end, strand specific libraries have remarkably improved transcriptome assembly performance and reliability as reflected by the relative statistics (see Table 1 in Article II).

The functional annotation of 267 differentially expressed genes that are shared between *C. compressirostris* and *C.tshokwe* has revealed marked differences in terms of metabolic pathways, classes of ion channels and regulatory networks that might be crucial for explaining the observed phenotypic differences between the electric organ and the skeletal muscle. Such differences suggest that the metabolic machinery of the electric organ could be mainly devoted to the production and turn-over of membrane structures. This form of specialization might be necessary in order to keep the peculiar anatomy of the stalk system. Additionally, a product of phospholipids metabolism, phosphatidylinositol 4,5-bisphosphate, is known for increasing the activity of ion channels (Suh & Hille 2005).

The up-regulation of the *atp1a2a* gene in the electric organ is explained by the fact that its product, the Na^+/K^+ ATP-ase, is fundamental for keeping the electrochemical gradient across the plasma membrane. Voltage-gated ion channels, on the other hand, are important for dissipating the electric potential generated by the ATP-ases and therefore for producing an EOD in response to an action potential. In the electric organ of the analyzed species, one gene coding for a voltage-gated sodium channel (*scn4aa*) is highly expressed in the electric organ, as already observed in other mormyrids and gymnotiform electric fishes. In fact, previous studies (Zakon *et al.* 2006) have demonstrated how the two paralogs of the voltage-gated sodium channelgene (*scn4aa*, *scn4ab*) are differentially expressed between skeletal muscle and electric organ in several unrelated species of electric fishes, whereas they are equally expressed in the muscle of *Danio rerio*. This finding suggests that paralog genes, which appeared after the teleost specific whole-genome duplication event, might have been involved in neo-functionalization processes through tissue-specific expression (Opazo *et al.* 2013).

Other over-expressed genes that increase cell excitability are the potassium channels *kcnq5a* and *kcnj9*. Both channels may act synergistically with sodium channels in order to promote the flux of cations across the electrocyte's plasma membrane.

The electric organ of mormyrid fishes still presents non-functional myofibrils; this vestigial presence is reflected in the transcriptome by the moderate expression of a few myosin and actin coding genes. However, most of the genes that are responsible for keeping a functional and organized sarcomere are down-regulated (e.g *tcap*), hence explaining the presence of dis-organized myofibrils in the electrocyte. Many genes involved in calcium compartmentalization (e.g. *atp2a1*, *atp2a2*, *casq1a*) and release (e.g. *ryr1*, *stac3*, *jph1*) are also down-regulated (Figure 1b). These results suggest that a key step in the evolution of electrogenesis might have been the disabling of the excitation-contraction pathway in the electrocytes.

The great number of transcription factors and signal transducers, identified as differentially expressed, suggests the important role played by regulatory networks in determining phenotypic differences between EO and SM. For instance the two basic Helix Loop Helix (bHLH) transcription factors *hey1* and *hes6*, in co-operation with *her6*, are known to negatively regulate the expression of myogenic factors in several model organisms (Figure 1a).

8.1.3 Future perspectives

The analysis of differentially expressed genes between skeletal muscle and electric organ in two species of African weakly-fishes suggests that: i) the loss of the contractile activity and the decoupling of the excitation-contraction processes are reflected by the down-regulation of the corresponding genes in the electric organ; ii) the metabolic activity of the EO might be specialized towards the production and turn-over of membrane structures; iii) several ion channels are highly expressed in the EO in order to increase excitability; iv) several myogenic factors might be down-regulated by transcription repressors in the EO.

Future transcriptomic analyses should investigate how these differences arise during development by performing RNA-Seq experiments at serial ontogenetic stages. In addition, the findings made in this study need to be experimentally validated *in vivo* by: i) identifying the patterns of tissue specific gene expression using techniques like fluorescent *in situ* hybridization, and ii) using reverse genetics approaches, like gene silencing, for inferring the function of candidate genes.

This study was primarily focused on investigating transcriptome-wide differences of gene expression level between different tissues within the same species. Further analyses, however, should investigate patterns of gene expression within the same organ across different species using recently derived models of gene expression evolution (Rohlfs *et al.* 2014). In fact, given the wide variability in terms of EOD characteristics across species one would expect greater inter-specific variance at the expression levels for the electric organs than for the skeletal muscle.

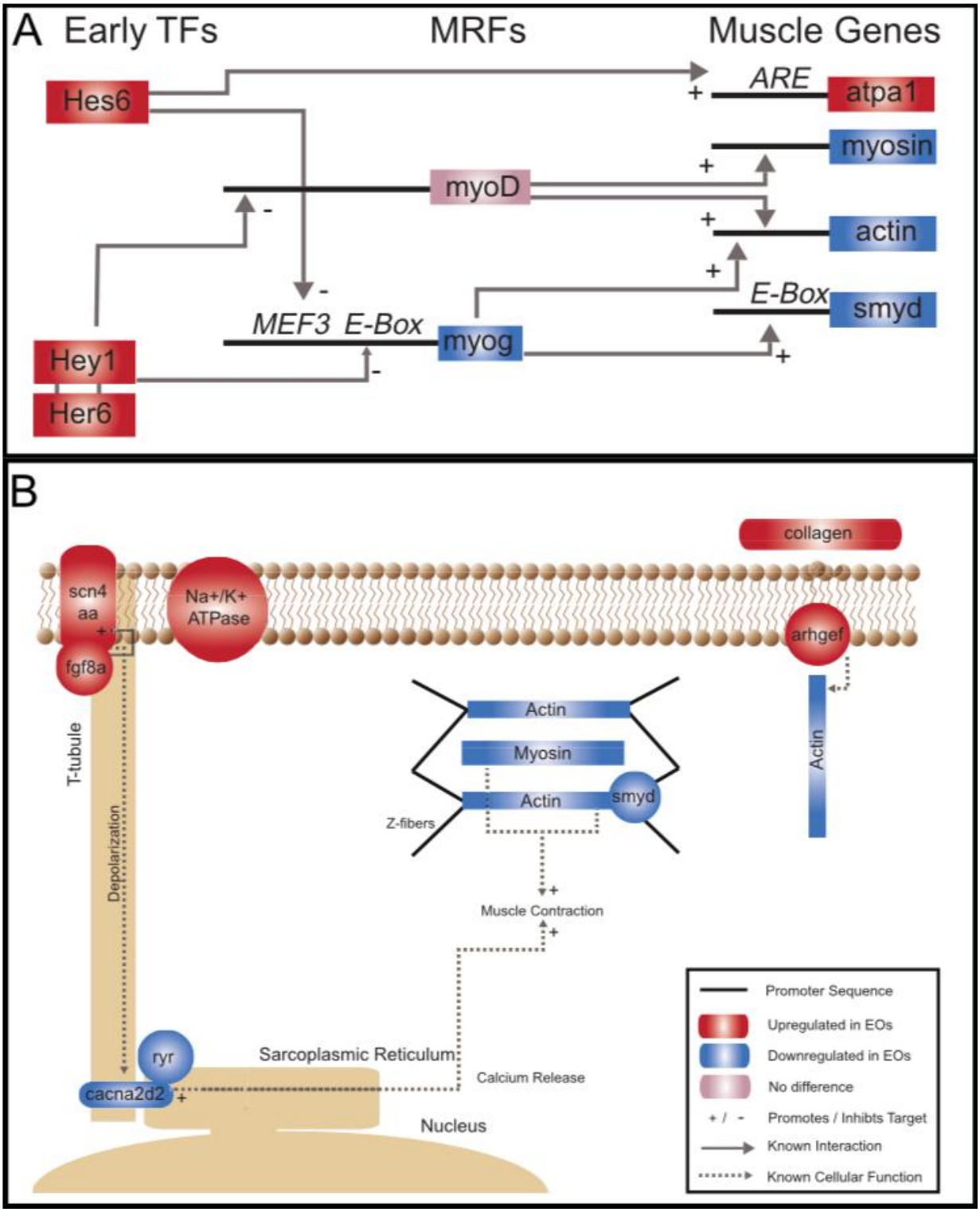


Figure 1 A) Interaction among the identified transcription factors and sarcomeric proteins. Early transcription factors (TFs) influence the expression of muscle regulatory factors (MRFs), which ultimately lead to the expression of muscle-specific effector genes. **B)** Picture showing the functioning of the excitation-contraction pathway in skeletal muscle. The down-regulation of genes involved in the calcium cycle determines the decoupling of the excitation and contraction mechanisms in the electric organ of *Campylomormyrus*.

8.2 *Campylomormyrus* species tree estimation and delimitation

The application of coalescent-based species tree estimation methods to *Campylomormyrus* has revealed the existence of two main species assemblages; one formed by: *C. compressirostris*, *C. curvirostris*, *C. rhynchophorus*, *C. numenius*, *C. tshokwe*, and *C. sp. nov.*; the other represented by all the members of the so-called "*alces*" complex (*C. alces*, *C. elephas*, and *C. chrystyi*). The basal position of *C. tamandua* confirms the results obtained by previous studies (Feulner *et al.* 2007). Species delimitation is statistically supported only for those taxa displaying marked differences in terms of morphology and EOD characteristics. An emblematic case is that of the two sister species *C. compressirostris* and *C. curvirostris*, which are characterized by prominently different snout morphologies and EOD shapes. In fact, although such evident differences, none of the previous molecular phylogenetic analyses was able to disentangle the relationships between these two taxa. On the other hand, the joint application of DNA sequence markers under a coalescent framework, and the use of microsatellites as an independent line of evidence have allowed to resolve their phylogenetic relationships. Additionally, the results of this study indicate that *C. sp. nov*. can be effectively considered as a distinct, yet undescribed species and confirm the results of the morphometric analysis conducted by Feulner *et al.* (2007). Its pale coloration and substantially reduced eyes suggest that this new species might be adapted to live in deep, dark waters (Figure 2).

Figure 2 *C. sp. nov.*: a yet undescribed species of *Campylomormyrus*.

All the putative cryptic species investigated in this study, which could be distinguished only by features of their electric discharge, did not obtain enough statistical support for being considered as distinguished species. There are mainly two explanations for this finding: i) the different EODs

observed are the result of environmental effects –i.e. phenotypic plasticity, or ii) these populations might be experiencing incipient sympatric speciation, driven by directional selection acting on few loci, and the markers used in this study may not be sufficient to capture any signature of reproductive isolation (Noor & Feder 2006; Nosil & Schluter 2011).

In conclusion, species-tree inference and species delimitation methods based on the multispecies coalescent model are very valid tools of investigation. Nevertheless, violations of their underlying assumptions may introduce inference errors. Previous simulation studies have demonstrated the robustness of the model against certain kinds of violations (e.g. gene flow among species; Ence & Carstens 2011; Zhang *et al.* 2011; Heled *et al.* 2013). However, when working with empirical data the consistency between the data and the model is not aprioristically known; for this reason the application of methods that can assess such fit is of vital importance. The usage of PPS model-checking techniques on the *BEAST species-tree reveals the coherence between the 5 markers developed for this study and the model, this does not necessarily mean that the topological conflicts among gene trees are the pure result of ILS (one of the main assumptions of the multispecies coalescent model), but that, if any other biological process (e.g. natural selection, hybridization, gene duplication) has shaped the present pattern of observed data, it does not drastically influence the outcome of an analysis based on such model.

8.2.1 Future perspectives

In order to increase the level of resolution among putatively cryptic species a considerably high amount of markers coming from unlinked loci should be utilized. The development of recent NGS techniques like RAD-Sequencing (Baird *et al.* 2008) would allow to develop enough Single Nucleotide Polymorphism (SNP) markers for achieving this task.

All the specimens analysed in the present study come from a restricted portion of the Congo River in the area of Kinshasa/Brazzaville. The Congo Basin, however, has a surface of 3.7 million square kilometres; therefore, future studies should investigate to what extent geographically distant populations are related to each other.

9 References

Alves-Gomes J, Hopkins CD (1997) Molecular Insights into the Phylogeny of Mormyriform Fishes and the Evolution of Their Electric Organs. *Brain, Behavior and Evolution*, **49**, 343–351.

Arnegard ME, McIntyre PB, Harmon LJ *et al.* (2010) Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *The American naturalist*, **176**, 335–56.

Baird N, Etter P, Atwood T, Currey M (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, **3**, e3376.

Bass A (1986) Electric organs revisited: evolution of a vertebrate communication and orientation organ. In: *Electroreception.* (eds Bullock TH, Heiligenberg W), pp. 13–70. Wiley, New York.

Bratton BO, Kramer B (1989) Patterns of the electric organ discharge during courtship and spawning in the mormyrid fish, Pollimyrus isidori. *Behavioral Ecology and Sociobiology*, **24**, 349–368.

Bullock T, Fay R, Hopkins C, Popper A (2005) *Electroreception.* Springer, New York.

Crawford JD (1991) Sex Recognition by Electric Cues in a Sound-Producing Mormyrid Fish, *Pollimyrus isidori*. *Brain, Behavior and Evolution*, **38**, 20–38.

Darwin C (1859) *On the origins of species by means of natural selection*. Murray, London.

Denizot JP, Kirschbaum F, Westby GWM, Tsuji S (1982) On the development of the adult electric organ in the mormyrid fish Pollimyrus isidori (with special focus on the innervation). *Journal of Neurocytology*, **11**, 913–934.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Von der Emde G (1999) Active electrolocation of objects in weakly electric fish. *J. Exp. Biol.*, **202**, 1205– 1215.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular ecology resources*, **11**, 473–80.

Feulner PGD, Kirschbaum F, Mamonekene V, Ketmaier V, Tiedemann R (2007) Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: Campylomormyrus): a combined molecular and morphological approach. *Journal of Evolutionary Biology*, **20**, 403–414.

Feulner PGD, Kirschbaum F, Schugardt C, Ketmaier V, Tiedemann R (2006) Electrophysiological and molecular genetic evidence for sympatrically occuring cryptic species in African weakly electric fishes (Teleostei: Mormyridae: *Molecular phylogenetics and evolution*, **39**, 198–208.

Feulner PGD, Kirschbaum F, Tiedemann R (2005) Eighteen microsatellite loci for endemic African weakly electric fish (Campylomormyrus, Mormyridae) and their cross species applicability among related taxa. *Molecular Ecology Notes*, **5**, 446–448.

Feulner P, Kirschbaum F, Tiedemann R (2008) Adaptive radiation in the Congo River: An ecological speciation scenario for African weakly electric fish (Teleostei; Mormyridae; Campylomormyrus). *Journal of Physiology-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. *Biology Letters*, **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. *Communicative & Integrative Biology*, **2**, 329–331.

Fujita MK, Leaché AD, Burbrink FT, McGuire J a, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology & evolution*, **27**, 480–8.

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. *The Journal of experimental biology*, **215**, 2479–94.

Gallant JR, Traeger LL, Volkening JD *et al.* (2014) Genomic basis for the convergent evolution of electric organs. *Science*, **344**, 1522–1525.

Gebhardt K, Böhme M, Von Der Emde G (2012) Electrocommunication behaviour during social interactions in two species of pulse-type weakly electric fishes (Mormyridae). *Journal of Fish Biology*, **81**, 2235–2254.

Heled J, Bryant D, Drummond AJ (2013) Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC evolutionary biology*, **13**, 44.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, **27**, 570–80.

Hopkins CD (2009) Electrical Perception and Communication C. In: *Encyclopedia of Neuroscience* (ed Squire LR), pp. 813–831. Academic Press., Oxford.

Hoskin CJ, Higgie M, McDonald KR, Moritz C (2005) Reinforcement drives rapid allopatric speciation. *Nature*, **437**, 1353–1356.

Irschick DJ, Vitt LJ, Zani PA, Losos JB (1997) A comparison of evolutionary radiations in mainland and Caribbean Anolis lizards. *Ecology*, **78**, 2191–2203.

KELLAWAY P (1946) The part played by electric fish in the early history of bioelectricity and electrotherapy. *Bulletin of the history of medicine*, **20**, 112–37.

Kirschbaum F (1983) Myogenic electric organ precedes the neurogenic organ in apteronotid fish. *Naturwissenschaften*, **70**, 205–207.

Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic biology*, **56**, 17–24.

Lavoué S, Miya M, Arnegard ME *et al.* (2012) Comparable ages for the independent origins of electrogenesis in African and South American weakly electric fishes. (WJ Murphy, Ed,). *PloS one*, **7**, e36287.

Lavoué S, Sullivan JP (2004) Simultaneous analysis of five molecular markers provides a well-supported phylogenetic hypothesis for the living bony-tongue fishes (Osteoglossomorpha: Teleostei). *Molecular Phylogenetics and Evolution*, **33**, 171–185.

Lissmann HW (1958) On the Function and Evolution of Electric Organs in Fish. *J. Exp. Biol.*, **35**, 156–191.

Lissmann HW, Machin KE (1958) The Mechanism of Object Location in Gymnarchus Niloticus and Similar Fish. *J. Exp. Biol.*, **35**, 451–486.

Losos JB (2010) Adaptive Radiation, Ecological Opportunity, and Evolutionary Determinism. *The American Naturalist*, **175**, 623–639.

Moller P (1995) Electric organs. In: *Electric Fishes: History and Behaviour* (ed Moller P), pp. 385–402. Chapman & Hall, London.

Near TJ, Eytan RI, Dornburg A *et al.* (2012) Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 13698–703.

Noor M, Feder J (2006) Speciation genetics: evolving approaches. *Nature Reviews Genetics*.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.

Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution*, **26**, 160–167.

Opazo JC, Butts GT, Nery MF, Storz JF, Hoffmann FG (2013) Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Molecular Biology and Evolution*, **30**, 140–153.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular biology and evolution*, **5**, 568–83.

Panhuis TM, Butlin R, Zuk M, Tregenza T (2001) Sexual selection and speciation. *Trends in Ecology and Evolution*, **16**, 364–371.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: The role of peripheral isolation. *Molecular Ecology*, **14**, 2943–2957.

Poll M, Gosse J, Orts S (1982) Le genre Campylomormyrus Bleeker, 1874, étude systematique et description d'une espece nouvelle (Pisces, Mormyridae). *Bulletin de l'Institut Royal des Sciences de Belgique Biologie*, **54**, 1–34.

Rohlfs R V, Harrigan P, Nielsen R (2014) Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Molecular biology and evolution*, **31**, 201– 11.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

Shadel GS, Clayton DA (1997) MITOCHONDRIAL DNA MAINTENANCE IN VERTEBRATES. *Annual Review of Biochemistry*, **66**, 409–435.

Smith JM (1966) Sympatric Speciation. *The American Naturalist*, **100**, 637.

Suh B-C, Hille B (2005) Regulation of ion channels by phosphatidylinositol 4,5-bisphosphate. *Current opinion in neurobiology*, **15**, 370–8.

Sullivan JP, Lavoué S, Hopkins CD (2000) Molecular Systematics of the African Electric Fishes (MormyroideD 7HOHRVWHLDQGD0RGHOIRUWKH(YROXWLRQRIWKHLU(OHFWULF2UJDQV*The Journal of experimental biology*, **683**, 665–683.

Szabo T (1960) Development of the Electric Organ of Mormyridae. *Nature*, **188**, 760–762.

Taverne L (1972) *Osteologie des genres Mormyrus Linne, Mormyrops Mueller, Hyperopisus Gill, Isichthys *LOO0\RP\UXV%RXOHQJHU6WRPDWRUKLQXV%RXOHQJHUHW*\PQDUFKXV&XYLHU FRQVLGHUDWLRQVJHQHUDOHV sur la systematique des poissons de l'ordre des Mormyriformes*. MRAC, Tervuren.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.

Zakon HH, Lu Y, Zwickl DJ, Hillis DM (2006) Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 3675–80.

Zhang C, Zhang D -X, Zhu T, Yang Z (2011) Evaluation of a bayesian coalescent method of species delimitation. *Systematic biology*, **60**, 747–61.

10 Acknowledgements

First of all, I would like to thank Prof. Dr. Ralph Tiedemann for giving me to opportunity to complete my doctoral studies in his lab on such an interesting and yet challenging topic. His constant encouragement and support have been really appreciated.

I also want to thank Prof. Dr. Frank Kirschbaum for sharing part of his huge knowledge on weakly electric fish with me. It was a great honor to work side by side in the field.

I am extremely grateful to Dr. Victor Mamonekene for the invaluable logistic and scientific support during my field trip in Congo.

A special thank goes to all my colleagues at the unit of evolutionary biology/systematic zoology for the very nice and relaxed working atmosphere. In particular I wish to thank the colleagues who have worked, or are currently working, on the same project as mine: Dr. Christiane Paul, Linh Nguyen and Rebecca Nagel, for sharing the effort of working with such a demanding group of fishes.

I would like to take this opportunity to thank all the members of the GENART network for supporting my research, and in particular Dr. Christoph Dieterich, Dr. Frieder Mayer and Isabelle Waurick.

I like to thank the group of bioinformatics at the University of Potsdam for providing all the "computational power" needed for my analyses.

Financial support was provided by the University of Potsdam and the Leibniz-SAW-project GENART.

Finally, I want to thank Elisa and my family for having continuously backed me and for their love and comprehension.