

RESEARCH ARTICLE

Open Access



Molecular evolution of globin genes in Gymnotiform electric fishes: relation to hypoxia tolerance

Ran Tian^{1,2}, Mauricio Losilla⁴, Ying Lu^{2,3}, Guang Yang^{1*} and Harold Zakon^{2,3*}

Abstract

Background: Nocturnally active gymnotiform weakly electric fish generate electric signals for communication and navigation, which can be energetically taxing. These fish mainly inhabit the Amazon basin, where some species prefer well-oxygenated waters and others live in oxygen-poor, stagnant habitats. The latter species show morphological, physiological, and behavioral adaptations for hypoxia-tolerance. However, there have been no studies of hypoxia tolerance on the molecular level. Globins are classic respiratory proteins. They function principally in oxygen-binding and -delivery in various tissues and organs. Here, we investigate the molecular evolution of alpha and beta hemoglobins, myoglobin, and neuroglobin in 12 gymnotiforms compared with other teleost fish.

Results: The present study identified positively selected sites (PSS) on hemoglobin (Hb) and myoglobin (Mb) genes using different maximum likelihood (ML) methods; some PSS fall in structurally important protein regions. This evidence for the positive selection of globin genes suggests that the adaptive evolution of these genes has helped to enhance the capacity for oxygen storage and transport. Interestingly, a substitution of a Cys at a key site in the obligate air-breathing electric eel (*Electrophorus electricus*) is predicted to enhance oxygen storage of Mb and contribute to NO delivery during hypoxia. A parallel Cys substitution was also noted in an air-breathing African electric fish (*Gymnarchus niloticus*). Moreover, the expected pattern under normoxic conditions of high expression of myoglobin in heart and neuroglobin in the brain in two hypoxia-tolerant species suggests that the main effect of selection on these globin genes is on their sequence rather than their basal expression patterns.

Conclusion: Results indicate a clear signature of positive selection in the globin genes of most hypoxia-tolerant gymnotiform fishes, which are obligate or facultative air breathers. These findings highlight the critical role of globin genes in hypoxia tolerance evolution of Gymnotiform electric fishes.

Keywords: Gymnotiformes, Hypoxia tolerance, Adaptive evolution, Globin, Positive selection

Background

Teleosts of the nocturnally-active neotropical clade Gymnotiformes produce and detect weak electric signals for the purposes of electrolocation and communication [1]. Variation in the patterning or frequency of the electric organ discharge (EOD) plays a vital role in electrical communication during behaviors such as aggression, courtship, and mating [2]. EODs are classified as wave-

or pulse-type. Wave-type EODs are formed by regularly-emitted pulses where the pulse duration is approximately equal to the inter-pulse interval thereby approximating a sine wave. Pulse-type EODs are emitted irregularly and the EOD pulse duration is short [3]. The order Gymnotiformes comprises six families: Hypopomidae, Rhamphichthyidae and Gymnotidae, which are all pulse-type, and Sternopygidae, Eigenmannidae and Apteronotidae, which are wave-type [4].

In most species EOs are composed of cells, called electrocytes, derived from muscle tissue. The EO is directly controlled by the nervous system, which commands the electrocytes to fire [3]. Electrocytes are large cells capable of generating large ionic currents, especially sodium

* Correspondence: gyang@njnu.edu.cn; h.zakon@austin.utexas.edu

¹Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China

²Department of Integrative Biology, The University of Texas, Austin, TX 78759, USA

Full list of author information is available at the end of the article



(Na⁺) currents. In order to restore the ionic gradient for Na⁺ following action potentials, electrocytes have large amounts of the “sodium pump” Na⁺/K⁺ ATPase; this pump uses one molecule of ATP for every three Na⁺ ions pumped back out of the cell. Thus, the generation of electricity requires energy. Physiological studies show that performance-related costs of EOD generation may be surprisingly high, from 10 up to 30% of routine oxygen consumption [5]. Yet, surprisingly, the O₂ consumption of gymnotiforms does not differ from that of other similarly sized teleosts [6] suggesting that gymnotiforms have adaptations for energy efficiency.

Gymnotiforms have their highest diversity throughout the ecologically varied lowland aquatic habitats of the vast Amazonian floodplains [7]. Although most gymnotiforms inhabit well-oxygenated streams and rivers (including all wave-type and a few pulse-type species), several lineages (only pulse-type species) have additionally radiated within oxygen-poor, stagnant habitats. Furthermore, all species are in danger of being trapped in shrinking, hypoxic pools during the dry season. Thus, hypoxia poses additional physiological challenges, most importantly, a deficiency in oxygen to fuel oxidative metabolism. As a consequence, coping with hypoxia is a potentially daunting task for gymnotiforms.

Different gymnotiform species possess a variety of anatomical and physiological adaptations to cope with these energetic demands including some common to other fish living in hypoxic environments such as large gill surface area [8], the use of aquatic surface respiration (ASR), in which fish swim to the surface and take in water from the topmost heavily oxygenated layer of water over their gills, and various forms of air-breathing [9]. In addition, some species have gymnotiform-specific means of conserving energy such as decreasing EOD amplitude (which lowers Na⁺ influx and, therefore, the amount of ATP needed to power Na⁺/K⁺ ATPase) during the day when they are inactive, upon encountering hypoxic conditions, or when food is scarce [10–12].

Little is known, however, about molecular adaptations for oxygen delivery or usage. Therefore, in this study, we focused on the well-known oxygen carriers, the globins. Globins are the most widespread respiratory proteins, existing in fungi, plants, and animals [13, 14]. Globins are conserved metalloproteins that typically have seven alpha helices that form a pocket containing an oxygen-binding heme [15]. They have been investigated for over a century, and eight globin types have been identified in vertebrates including hemoglobin (Hb), myoglobin (Mb), and neuroglobin (Ngb). Hb and Mb are best known for their respiratory functions, which play critical roles in the maintenance of cellular oxygen supply in support of aerobic metabolism [16, 17].

Hb is a heterotetramer protein composed by two α - and two β -chains, which are encoded by the corresponding alpha (Hba) and beta (Hbb) globin gene family [18]. Moreover, it is responsible for facilitating O₂ from the respiratory system to the inner organs via the circulatory system [18]. Mb is a compact and highly soluble monomer protein containing one proto-heme, which is involved in the oxygen storage and transportation within heart and skeletal muscle cells [19], and has a higher oxygen affinity than Hb. Ngb is a monomer and an oxygen-carrying protein essentially restricted to neurons [20], which plays a key role in facilitated diffusion and local storage of O₂. Although its function is still uncertain, there is general agreement that Ngb is associated with mitochondria and thus oxidative metabolism, and serves a neuroprotective role during hypoxic stress [13]. Interestingly, Hb, Mb and Ngb have also been proposed to have roles in nitric oxide (NO) metabolism and the detoxification of reactive oxygen species during hypoxia [21].

We gathered coding sequences of Hba, Hbb, Mb, and Ngb genes from 12 gymnotiform species to: 1) test whether these globin genes have evolved adaptively (i.e.,—show signs of positive selection) during gymnotiform origins and evolution; 2) evaluate whether gymnotiforms that inhabit hypoxic/anoxic vs. well-oxygenated water display different patterns of molecular evolution and; 3) provide a more comprehensive picture of the hypoxia tolerance in gymnotiforms.

Methods

Taxon sampling, and primary treatments of data

A total of 12 individuals representing six families across Gymnotiformes were analyzed in our study (see Table 1). Some species live in habitats with persistent hypoxia or anoxia while the others are unable to tolerate severely hypoxic water (see Table 1). In order to attain a broad and balanced taxonomic coverage, we also obtained sequences from Siluriformes, Characiformes, Cypriniformes, Gonorynchiformes, and Clupeiformes (see Additional file 1: Table S1). The tree topology used for the analysis is depicted in Fig. 1 [4]. Nucleotide and protein sequences of species were downloaded from National Center for Biotechnology Information (NCBI: <http://www.ncbi.nlm.nih.gov/>), from NextGen-derived sequences available on a website hosted by the Electric Fish Genome Consortium (<http://efishgenomics.zoology.msu.edu/blast>), and/or amplified by Polymerase Chain Reaction (PCR). A complete list of non-gymnotiform specimens and accession numbers are in Additional file 1: Table S1.

Genomic DNA and tissue RNA isolation and sequencing

Genomic DNA was extracted using QIAGEN DNeasy Tissue Kit (QIAGEN, Inc.). The total RNA was isolated

Table 1 Characteristics of species used in this study

Family	Species	Hypoxia-tolerant	Air breather	EOD type	Positive selected genes
Gymnotidae	<i>Electrophorus electricus</i>	yes	obligate	Pulse	<i>Mb, Hba</i>
Gymnotidae	<i>Gymnotus cylindricus</i>	yes	facultative	Pulse	
Gymnotidae	<i>Gymnotus omarorum</i>	yes	facultative	Pulse	
Hypopomidae	<i>Brachyhypopomus gauderio</i>	yes	facultative	Pulse	<i>Mb, Hbb</i>
Hypopomidae	<i>Microsternarchus bilineatus</i>	no	no	Pulse	
Hypopomidae	<i>Steatogenys elegans</i>	no	no	Pulse	<i>Mb</i>
Rhamphichthyidae	<i>Gymnorhamphichthys sp.</i>	no	no	Pulse	
Rhamphichthyidae	<i>Rhamphichthys marmoratus</i>	no	no	Pulse	
Apterontidae	<i>Apteronotus albifrons</i>	no	no	Wave	
Apterontidae	<i>Parapteronotus hasemani</i>	no	no	Wave	
Sternopygidae	<i>Eigenmannia virescens</i>	mildly	no	Wave	<i>Hbb</i>
Sternopygidae	<i>Sternopygus macrurus</i>	no	no	Wave	<i>Mb, Ngb</i>

Note: Designation of hypoxia tolerance based on habitat choice and physiological tests of hypoxia tolerance [10, 11, 72]

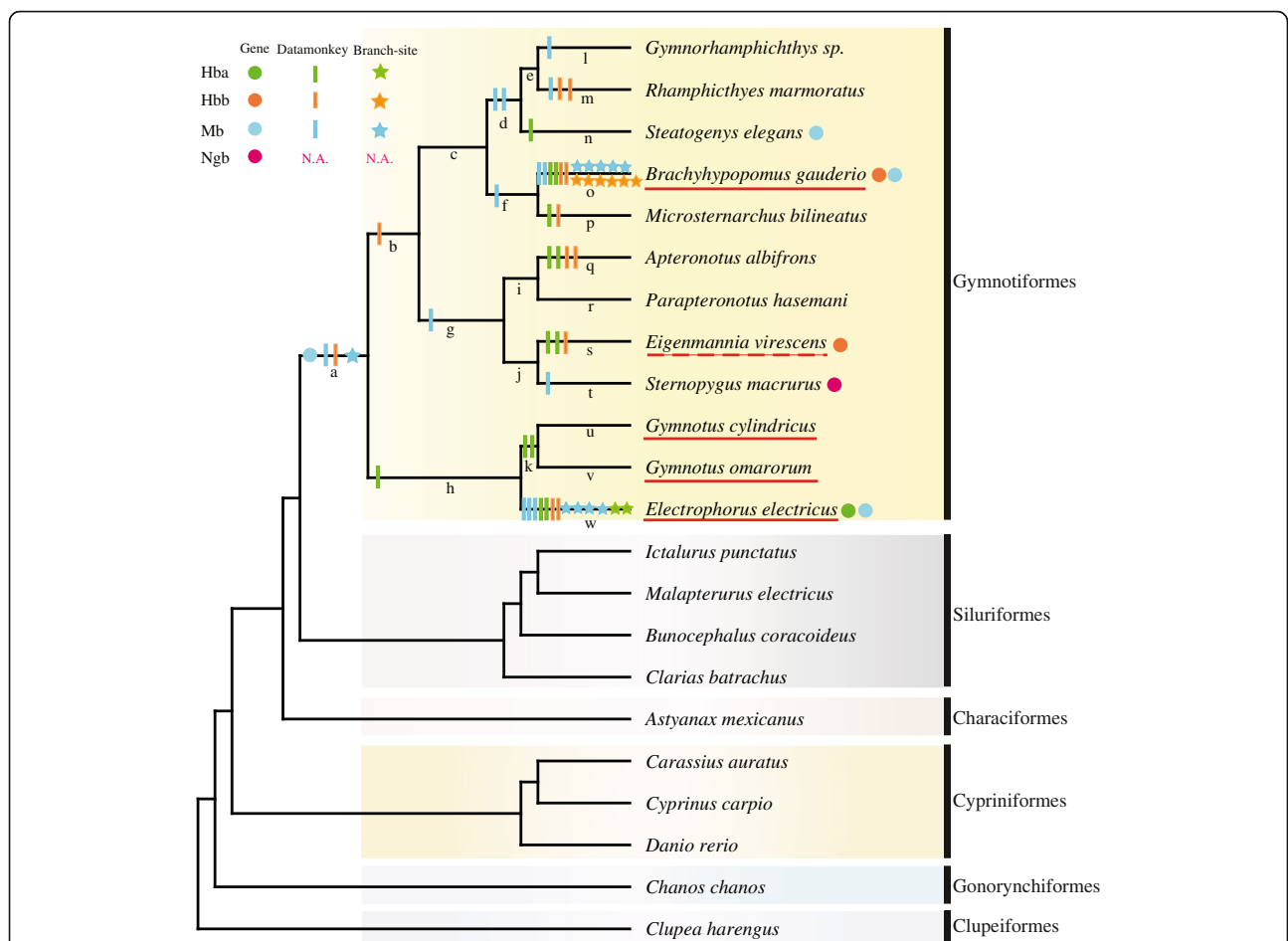


Fig. 1 Radical amino acid changes of positively selected sites are shown above the branches across gymnotiformes species tree from a-w. Circles indicate positively selected genes across branches. Bars represent radical amino acid changes of sites detected by Datamonkey. Stars indicate amino acid substitution identified by branch-site model. Species underlined are facultative (or obligate in the case of *E. electricus*) air breathers. Dotted species (*E. virescens*) is mildly hypoxia tolerant. Hba, Hbb, Mb and Ngb are colored with green, orange, blue and pink, separately

from skeletal muscle, electric organ and brain using the QIAGEN RNeasy Mini Kit (QIAGEN, Inc.), and used as a template for single-stranded (ss) cDNA synthesis using the SuperScript III reverse transcriptase protocol (Invitrogen). Degenerate and specific primers were designed using the homologous sequences from the genome of electric fishes by using Oligo [22]. The primers are listed in Additional file 1: Table S3. This ss cDNA (1:10 dilution) and genomic DNA were used as the template with degenerate primer pairs in the first nested PCR, and their products were templates in second nested PCR reactions with primer pairs. PCR amplifications were carried out using the following program: 95 °C 4 min, 35 cycles of 94 °C denaturation 1 min, 55–50 °C annealing 30 s, 72 °C extension 1 min, and 72 °C elongation for 15 min. PCR products were sequenced in both directions. The newly determined sequences were deposited in GenBank under the accessions KX827275-KX827283, KX833066-KX833070 and KX982253- KX982258.

Phylogenetic reconstruction

Sequences were assembled and primer sequences were trimmed using MacVector [23]. Nucleotide sequences were translated into amino acid sequences and aligned using MEGA 6.0 [24]. The alignments were manually inspected and edited by eye (see Additional file 2: Figure S1). The best fit model of nucleotide evolution was determined by jModeltest [25], and Maximum Likelihood (ML) trees was reconstructed by PhyML 3.0 [26].

Molecular evolutionary analysis

To determine whether adaptive evolution might have occurred on the globin genes of electric fishes, we used the PAML package [27], which uses a maximum-likelihood (ML) approach to calculate nonsynonymous to synonymous rate ratios ($\omega = dN/dS$). The ratios >1 , $=1$ and <1 indicate positive selection, neutrality and negative selection, respectively. For all genes, both the species tree (Fig. 1) and putative ML tree (Additional file 2: Figure S2) were separately used as the working topology in all the analyses.

We used a site model for positive selection at individual codons in electric fish samples for each gene, i.e., M8 and M8a [28]. Model M8a only allows codons to evolve neutrally or under purifying selection ($\omega < 1$), whereas M8 model includes a class of sites with $\omega > 1$. Amino acids under selection for model M8 were identified using a Bayes empirical Bayes approach (BEB), and we considered as candidates sites with a posterior probability $>80\%$ [29]. Then, we further employed a series of ML methods implemented in the Datamonkey web server (<http://www.datamonkey.org>), which has the advantage of improving the estimation of the dN/dS ratio by incorporating variation in the rate of synonymous substitution [30]. The single likelihood ancestor counting

(SLAC) model is a conservative test, which counts the synonymous and nonsynonymous changes at each codon position in a phylogeny. The Fixed-Effect Likelihood (FEL) calculates site-by-site dN/dS without assuming a prior distribution. The Random-Effect Likelihood (REL) assumes a prior distribution across sites. In addition, the Fast Unconstrained Bayesian Approximation (FUBAR) ensures robustness against model misspecification. Each module was run using the default cutoffs with $p = 0.2$ for SLAC and FEL, Bayes Factor = 50 for REL and posterior probability = 0.8 for FUBAR.

To test for possible heterogeneity of ω ratios along independent branches, we used the free-ratio model, which allows each branch to have a separate dN/dS . The null model is a very strict model called the one-ratio model (M0) that allows only a single ω ratio for all branches. We further executed branch-site tests to explore positive selection affected by a few sites along a specific branch [31]. We compared modified model A, which assumes four classes of sites, especially, allowing codons under positive selection along foreground lineage with $\omega_2 > 1$, to the null hypothesis, in which fixed $\omega_2 = 1$ is allowed based on branch-site model A. For all the analyses, the nested models were compared using a likelihood ratio test (LRT) with various degrees of freedom, and all analyses were run twice to ensure convergence. Branch-site REL analysis was also performed to determine whether specific-lineage is evolving under positive selection using a web-based implementation of HyPhy package (<http://www.datamonkey.org>), which is based on likelihood ratio tests that identify all lineages with a proportion of sites that are evolving with $dN/dS > 1$, and do not require partitioning lineages into foreground and background branches [32].

Each gene sequence alignment was also analyzed using the program TreeSAAP [33], which further supports PAML and Datamonkey at the protein physicochemical level. TreeSAAP compares the magnitude of physicochemical changes of non-synonymous residues across a phylogeny and identifies specific amino acid properties that have likely been affected by positive destabilizing selection during evolution. In this study, amino acid properties were considered to be under positive-destabilizing selection if positive selection was detected in radical magnitude ranges (6–8). The number of radical changes in the amino acid properties was used as a proxy for determining the strength of positive selection. More radical changes in amino acid properties might suggest adaptive evolution. The residues that had fewer than six amino acid property changes were categorized as type I sites, whereas those that had more than six were categorized as type II sites.

Structural analyses

To provide further insights into the underlying effects of these positively selected sites, we mapped them onto the

three-dimensional (3D) structure. The crystal structures of P11748 for Hba, P11749 for Hbb, and 2NRL for Mb were taken from the Protein Data Bank (<http://www.rcsb.org/pdb>). Postulated functional regions or residues were searched from Uniprot (<http://www.uniprot.org/uniprot>). Pymol [34] was used to produce the images of the three-dimensional models of corresponding gene.

Transcriptome assembly

Tissues were removed from a single *Brachyhyppopomus gauderio* that was previously housed under normoxic conditions in the laboratory and total RNA was extracted from brain, skeletal muscle, heart and electric organ. RNA was treated with RiboZero (Illumina, MRZH-11124) kit to remove ribosomal RNA, and cDNA libraries (200 bp, paired ends) were made. Libraries were sequenced on an Illumina HiSeq machine. We processed the raw reads with Trimmomatic v0.32 [35] for adapter removal (IlluminaClip: TruSeq3-PE.fa:2:30:10), quality trimming (SlidingWindow:4:5, Leading:5, Trailing:5) and size filtering (MinLen:25). These are Trinity's default settings, which are based on the work of MacManes [36]. We performed quality control of both raw and processed reads with FastQC v0.11.3 (<http://www.bioinformatics.babraham.ac.uk/>) (Table S6).

We combined the processed PE reads across organs, *in silico* normalized, and *de novo* assembled them with Trinity v2.2.0 [37, 38]. Then we used BUSCO v1.1b1 [39], along with BLAST+ v2.2.31 [40], HMMER v3.1b1 [41], and EMBOSS v6.5.7 [42] to measure transcriptome completeness, against the Vertebrates, Metazoans and Eukaryotes datasets. In all cases, the assembly displayed a large percentage of complete orthologs (Table S7).

We estimated levels of expression for each transcript and gene, per organ, using the Trinity-provided scripts `align_and_estimate_abundance.pl` and `abundance_estimates_to_matrix.pl`. We used both RSEM v1.2.19 [43] and kallisto v0.42.5 [44] quantification methods, which produced qualitatively very similar results. Reported abundances are TMM-normalized values calculated with the RSEM method. Only one Trinity gene blasted against the myoglobin gene sequence with an E-value of 0.00. This gene's per organ abundances are the ones reported.

The abundance estimation results suggested that very few genes accounted for a large fraction of gene expression. Upon inspection of said genes, many were related with rRNA, and therefore were expected to be depleted during the library preparation process. Although our *B. gauderio* transcriptome assembly meets our quality standards, we don't recommend future use of the RiboZero kit (which is designed for human, mouse and rat) when working with RNAseq from this taxon. Sequences for *B. gauderio* genes are available at: http://efishgenomics.integrativebiology.msu.edu/blast_search/.

Results

Phylogenetic analyses

We successfully amplified cDNAs for the Mb, Hba, Hbb and Ngb gene from eight Gymnotiforms species (see Additional file 1: Table S2). There are no insertion/deletion mutations or changes that result in stop codons in gene sequences, suggesting the presence of functional proteins in electric fishes. These were added to mRNA sequences previously obtained by NextGen sequencing from four other species for a total of 12 species.

We constructed phylogenetic trees using maximum likelihood (ML) performed by PhyML from nucleotide alignments of the four genes dataset. The Akaike Information Criteria (AIC) in jModeltest selected the GTR + G substitution model for all genes. Relationships of the gene trees largely reflected species relationships previously estimated with morphological data by Tagliacollo et al. [4] (see Additional file 2: Figure S2). For example, the phylogenetic tree placed Gymnotiformes and Siluriformes together, and they had a closer relationship with Characiformes than Cypriniformes in the Mb gene tree. However, the topology based on the genes still failed to resolve the relationships within Gymnotiformes, which is not surprising since this has been difficult to resolve even with large datasets. The bootstrap support values for these relationships were not high, which is most likely due to the short length of globin genes and small number of species that were used to reconstruct the topologies.

Adaptive molecular evolution

Considering that selection analyses using the species tree was basically the same as that using the gene trees, only the former analyses are shown here. First, we used a pair of site models (M8 vs. M8a) to address whether recurrent positive selection has acted on specific codons in globin genes. Likelihood ratio tests (LRTs) showed that a model that includes sites with $dN/dS > 1$ (M8) fits the data significantly better than a neutral model (M8a) for the Hbb gene, and one positively selected site was identified with high posterior probability (pp) (Hbb: 133, pp = 0.996) (see Additional file 1: Table S4). Because CodeML does not incorporate rate variation in synonymous sites (d_S), we therefore further analyzed the data using the SLAC, FEL, REL, and FUBAR model implemented in the Datamonkey website, which have the advantage of improving the estimation of the ω ratio incorporated variation in the rate of synonymous substitution (Pond and Frost 2005a). Similarly, four ML methods also detected sites under selection for Mb, Hba and Hbb genes, some of which coincide with the codons previously identified by M8 (see Table 2). Seven positively selective sites were identified using the SLAC method at significance level of 0.2 (Mb: 2, Hba: 3,

Table 2 Evidence of positive selection identified by different methods in Hyphy

Gene	SLAC (P > 0.8)	FEL (P > 0.8)	REL (PP > 50)	Fubar (PP > 0.8)	Branch-site REL
<i>Hba</i>	<u>35</u> (0.134), <u>73</u> (0.199), <u>109</u> (0.133)	<u>35</u> (0.058), <u>73</u> (0.077), <u>79</u> (0.111), <u>83</u> (0.132), <u>109</u> (0.073)	<u>35</u> (60.944), <u>73</u> (60.689), <u>83</u> (50.668), <u>109</u> (84.333)	<u>35</u> (0.909), <u>73</u> (0.840), <u>79</u> (0.940), <u>109</u> (0.824)	<i>E.electricus</i> : p = 0.007
<i>Hbb</i>	<u>123</u> (0.197), 133 (0.154)	<u>15</u> (0.093), <u>45</u> (0.146), <u>81</u> (0.127), <u>106</u> (0.164), <u>112</u> (0.162), <u>123</u> (0.072), 133 (0.067), <u>144</u> (0.100)	<u>123</u> (60.003), 133 (1832.56)	<u>15</u> (0.818), <u>81</u> (0.874), <u>111</u> (0.899), <u>123</u> (0.854), 133 (0.979), <u>144</u> (0.921)	<i>B.gauderio</i> : p = 0.003; <i>E.virescens</i> : p = 0.037
<i>Mb</i>	<u>27</u> (0.097), <u>98</u> (0.096)	<u>27</u> (0.021), <u>92</u> (0.110), <u>98</u> (0.029), <u>108</u> (0.131)	<u>27</u> (337.517), <u>98</u> (94.269)	<u>27</u> (0.973), <u>44</u> (0.849), <u>92</u> (0.849), <u>98</u> (0.925)	<i>E.electricus</i> : p = 0.005 <i>B.gauderio</i> : p = 0.017

Note: Codons identified by more than one ML method are underlined. Sites also detected by M8 model are shown in bold. Site positions were relative to Zebrafish (*Danio rerio*) protein sequences, i.e., Mb (Q6VN46), Hba (Q90487), Hbb (Q90486), and Ngb (Q90YJ2) in UniProtKB

Hbb: 2). The FEL methods found 17 positively selected codons (Mb: 4, Hba: 5, Hbb: 8). Using the REL method, 8 codons were identified under positive selection at significant level of Bayes factor > 50 (Mb: 2, Hba: 4, Hbb: 2). Furthermore, FUBAR also identified 14 sites under diversifying selection with a posterior probability > 0.8 (Mb: 4, Hba: 4, Hbb: 6) (see Table 2). It is generally accepted that a positively selected site is more reliable if it can be supported by two or more different methods. Of these putative positively selected sites, 13 codons (Mb: 27, 92, 98; Hba: 35, 73, 79, 83, 109; Hbb: 15, 81, 123, 133, 144) were detected by more than one ML methods, which are robust candidates for sites under selection; especially, site 133 in Hbb was determined to have undergone positive selection by all methods (see Table 2).

To test for positive selection along particular branches, we used the one-ratio model (M0) that allows only a single ω ratio for all branches. The ω ratio estimated by M0 is significantly less than 1 (from 0.075 to 0.296) (see

Additional file 1: Table S5), suggesting that strong purifying selection plays a central role in the evolution of globin genes in electric fishes. The codeml-free ratio branch model of PAML estimated independent ω along all branches of the phylogeny, which was significantly better than the one-ratio model ($p < 0.001$, see Additional file 1: Table S5) for Mb, Hba and Hbb, suggesting heterogeneous selective pressures on different lineages. To investigate whether the evidence for positive selection was restricted to individual codons along some specific lineages, branch-site models were then performed. Foreground branches for tests of positive selection were selected on clades of Gymnotiforms. The LRT tests showed that four Mb branches, one Hba branch, one Hbb branches and one Ngb branch exhibited significant ($P < 0.05$) evidence of positive selection over background branches, separately (see Table 3). Further evidence of positive selection was also detected by Branch-site REL. The results show that lineages leading to *Brachyhypopomus gauderio* (branch o, $p = 0.017$) and *Electrophorus electricus* (branch w, $p =$

Table 3 Selective pressure analyses of Gymnitiformes by the Branch-Site Model

Gene	Branch	Model	Ln	$2\Delta Ln$	p	Parameter	Positive selected sites
<i>Hba</i>	<i>E.electricus</i>	Ma	3039.610			$\omega_0 = 0.129, \omega_1 = 1, \omega_2 = 6.251$	18-0.995**, 46-0.990*
		Ma0	3042.305	5.390	0.020	$\omega_0 = 0.129, \omega_1 = 1, \omega_2 = 1$	
<i>Hbb</i>	<i>B.gauderio</i>	Ma	3177.724			$\omega_0 = 0.145, \omega_1 = 1, \omega_2 = 34.694$	20-0.955*, 29-0.973*, 113-0.989*, 114-0.912, 132-0.981*, 133-0.888
		Ma0	3187.339	19.230	<0.001	$\omega_0 = 0.139, \omega_1 = 1, \omega_2 = 1$	
<i>Mb</i>	<i>B.gauderio</i>	Ma	5191.675			$\omega_0 = 0.137, \omega_1 = 1, \omega_2 = 15.025$	9-0.987*, 18-0.964*, 98-0.940, 109-0.990**, 122-0.994**, 134-0.992**, 138-0.911
		Ma0	5199.405	15.46	<0.001	$\omega_0 = 0.136, \omega_1 = 1, \omega_2 = 1$	
	<i>E.electricus</i>	Ma	5191.521			$\omega_0 = 0.136, \omega_1 = 1, \omega_2 = 381.189$	37-0.998**, 106-0.982*, 108-0.960*, 109-0.993**, 117-0.944
		Ma0	5195.127	7.212	0.007	$\omega_0 = 0.132, \omega_1 = 1, \omega_2 = 1$	
	<i>S.elegans</i>	Ma	5201.909			$\omega_0 = 0.142, \omega_1 = 1, \omega_2 = 999$	38-0.933
		Ma0	5204.687	5.556	0.018	$\omega_0 = 0.141, \omega_1 = 1, \omega_2 = 1$	
	Last common ancestor of gymnotiform	Ma	5199.339			$\omega_0 = 0.143, \omega_1 = 1, \omega_2 = 60.065$	12-0.986*
Ma0		5204.040	9.402	0.002	$\omega_0 = 0.141, \omega_1 = 1, \omega_2 = 1$		
<i>Ngb</i>	<i>S.macrurus</i>	Ma	2223.853			$\omega_0 = 0.063, \omega_1 = 1, \omega_2 = 189.298$	150-0.989*
		Ma0	2226.360	5.014	0.025	$\omega_0 = 0.063, \omega_1 = 1, \omega_2 = 1$	

Note: *: posterior probability (pp) >95%; **: pp >99%

0.005) were under positive selection for the Mb gene. Additionally, this model also confirms episodic selection along the same lineages (Mb: branch o, $p = 0.017$ and branch w, $p = 0.005$; Hba: branch w, $p = 0.007$; Hbb: branch o, $p = 0.003$) reported by the branch-site model in PAML, but also identifies an additional lineage—the *Eigenmannia virescens* (branch s, $p = 0.037$) lineage for Hbb gene (see Table 2). To summarize the above results from the analyses using different methods, all four globin genes were identified as having undergone positive selection in Gymnotiforms.

Although Codeml and Datamonkey estimate the influence of natural selection at the codon level, selection for change in amino acid physicochemical properties was also analyzed by TreeSAAP, which further support the ML methods results at a complementary protein-level. Overall, 33 residues in four globin genes were subjected to positive selection supported by two or more ML methods (Datamonkey, site model and Branch-site model), and 78.8% (26/33) (Hba: 6, Hbb: 9, Mb: 11) were also detected by TreeSAAP as under positive selection at the physicochemical level (see Table 4). Nine of these sites were Type II class (more than or equal to six radical changes in amino acid properties). Furthermore, the radical amino acid changes in the 26 positively selected codons were scattered throughout most of the Gymnotiform phylogeny (see Fig. 1). It is interesting to note that high levels of positive selection were found in “hypoxic water” species compared with “well-oxygenated water” species; that is, 15 radical amino acid sites (Hba: 2, Hbb: 7, Mb: 6) were detected in branch o; 13 radical amino acid sites (Hba: 4, Hbb: 2, Mb: 7) were identified in branch w; and 3 radical amino acid sites (Hba: 2, Hbb: 1) were found in branch s (see Fig. 1).

Structural links to protein function

To gain insight into the functional significance of the putatively selected sites, we mapped all the radical amino acid sites onto 3D structures. We found that several sites fall in or immediately adjacent to the functional regions or residues (see Fig. 2). For example, site 83 in Hba gene was close to the heme proximal ligand residue (site 89 in Zebrafish). Again, site 144 in Hbb was adjacent to the β -147 histidine. For the Mb gene, sites 9, 12, 27, 37 and 38 were located in the globin region. 42.3% (11/26) positively selected sites with radical changes were localized in residues postulated to affect function.

Expression profile of Mb and Ngb in *E. Electricus*

We wished to compare expression patterns under normoxic conditions of *E. electricus* and *B. gauderio* Mb and Ngb with other teleosts. We did not measure levels of hemoglobins in these tissues because hemoglobins are only expressed in erythrocytes; any signal from hemoglobins would be from residual red cells left in the tissues

rather than expression in the tissues themselves. As a result of earlier work [45, 46], we had access to tissue transcriptomes of *E. electricus* and examined levels of Mb and Ngb expression (see Fig. 3). Mb is highly expressed in the heart and less so in muscle. Mb is also expressed in some non-muscle-derived tissues such as brain, spinal cord, and kidney. Surprisingly, it is only negligibly expressed in the EOs (*E. electricus* has two weak EOs for communication and navigation and one strong EO for shocking prey). Ngb is expressed in the brain, spinal cord, and kidney, and at negligible levels elsewhere (see Fig. 4).

Transcriptomes were made from some of the same tissues (brain, heart, skeletal muscle, EO) of *B. gauderio* (see Additional file 1: Table S6 and S7; Additional file 3). The overall pattern of Mb expression is similar to *E. electricus*. That is, expression is highest in the heart, also observed in brain, and low in muscle and EO. Ngb levels were too low to measure accurately.

Discussion

Pervasive adaptive evolution of gymnotiform globin genes

Recent studies have shown strong evidence for positive selection in hypoxia tolerant species, like yak [47], hummingbirds [48], and cetaceans [49], when compared with their hypoxia intolerant relatives, as well as hypoxia-tolerant populations of humans such as Tibetans [50].

In this study, we surveyed all four primary oxygen-carrier globin genes in gymnotiforms for signs of positive selection and to assess whether hypoxia tolerance has influenced the evolution of these genes. Our analyses provide strong evidence that globin genes have been subjected to positive selection during gymnotiform evolution. First, neutral models of evolution were rejected for Hbb genes, and more than two ML methods identified specific codons with a high probability of being under selection for Hba, Hbb and Mb genes. Second, adaptive evolution was further supported by evidence of radical changes in positively selected amino acids in gymnotiform globins. Again, 34.6% (9/26) belong to the Type II class, suggesting robust positive selection. Finally, several of the putatively selected sites fall in, or close to, regions important for function based on structural information.

Positively selected sites are scattered pervasively along lineages of gymnotiform phylogeny (see Table 2 and Fig. 1), suggesting the contribution of the respiratory proteins for oxygen storage and transportation during adaptation to expensive oxygen consumption in gymnotiforms. Moreover, a signal of positive selection was also detected in the lineage leading to the common ancestor of gymnotiforms. This lineage represents the early evolutionary history of the gymnotiform's evolved EOs, during which the gymnotiforms were faced with the challenges

Table 4 Radical amino acid sites under positive selection detected by Datamonkey, Branch-site model and TreeSAAP Simultaneously

Gene	Position ^a	Clade ^b	AA Changes	Radical Changes in AA Properties ^c	Total	
<i>Hba</i>	18	w	I-A	$\rho K', R_{ar}$	2	
	<u>35</u>	k	T-A	P_{α}	1	
		s	T-V	N_s, R_{α}	2	
		n	T-P	K°, H_t	2	
		o	T-I	$N_s, B_r, \rho K', R_{ar}, H_{pr}, H_t$	6	
	46	w	H-S	F, E_t	2	
	<u>73</u>	h, s	S-A	$P_{\alpha}, P_{\alpha}, P$	3	
		w	A-G	$P_{\alpha}, P_{\alpha}, P$	3	
	<u>79</u>	o, w	A-T	P_{α}	1	
		p, q	A-C	$P_{\alpha}, N_s, c, K^{\circ}, \rho K', \alpha_n$	6	
	<u>83</u>	k, q	A-S	$P_{\alpha}, P_{\alpha}, P$	3	
	<i>Hbb</i>	<u>15</u>	p	M-I	$\rho K'$	1
		20	o	S-H	F	1
29		o	L-I	$\rho K'$	1	
<u>81</u>		s	N-A	$P_{\alpha}, P_{\alpha}, P$	3	
113		o	V-T	N_s, R_{α}	2	
114		o	C-T	N_s, B_r, c	3	
<u>123</u>		m, w	V-G	$P_{\beta}, B_r, P_{\alpha}, F, R_{ar}, P$	6	
		o, q	V-A	P_{β}	1	
132		o	W-I	$\rho K'$	1	
<u>133</u>		a	Q-C	N_s, B_r, c, h, p, E_t	6	
		h	C-S	N_s, B_r, μ	3	
		w	S-A	$P_{\alpha}, P_{\alpha}, P$	3	
		q	C-Q	N_s, B_r, c, h, p, E_t	6	
	m	C-T	N_s, B_r, c	3		
	o	C-E	$P_{\alpha}, N_s, B_r, c, h, E_t, P_r, p, \alpha_c, H_{pr}, E_t$	11		
<i>Mb</i>	9	o	K-A	$E_t, B_r, K^{\circ}, h, E_{sm}$	5	
	12	a	G-P	B_r, α_c, H_t	3	
	<u>27</u>	w	T-G	B_r	1	
		l	T-V	N_s, R_{α}	2	
		o	S-W	$B_r, R_F, P_{\alpha}, C_{\alpha}, E_t, F, M_w, M_{wr}, V^{\circ}, u, R_{ar}, H_t$	12	
	37	w	Q-I	$N_s, B_r, \rho K', h, F, p, E_t, E_t, R_{ar}, H_{pr}, H_t$	11	
	<u>92</u>	w, f	K-Q	Ph_i	1	
		g	K-I	$N_s, P_{\beta}, B_r, R_F, \rho K', h, E_t, F, H_{ncr}, P_r, p, R_{ar}, H_{pr}, E_t$	14	
		t	I-V	$\rho K'$	1	
		d	K-T	Ph_i, E_t	2	
	<u>98</u>	a	N-A	$P_{\alpha}, P_{\alpha}, P$	3	
		w	A-I	$\rho K', R_{\alpha}$	2	
		d	A-G	$P_{\alpha}, P_{\alpha}, P$	3	

Table 4 Radical amino acid sites under positive selection detected by Datamonkey, Branch-site model and TreeSAAP Simultaneously (Continued)

	m	G-V	$P_{\beta}, B_r, P_{\alpha}, F, R_{ar}, P$	6
	o	A-N	$P_{\alpha}, P_{\alpha}, P$	3
106	w	V-C	$\rho K', C, R_{ar}, P$	4
108	w	V-C	$\rho K', C, R_{ar}, P$	4
109	w	K-V	$N_s, B_r, h, E_t, F, p, R_{ar}, H_{pr}, H_{ncr}, E_t$	10
	o	K-S	E_t	1
122	o	Q-A	K°, E_t	2
134	o	N-A	$P_{\alpha}, P_{\alpha}, P$	3

^aRelative to Zebrafish (*Danio rerio*) protein sequences, i.e., Mb (Q6VN46), Hba (Q90487), and Hbb (Q90486) in UniProtKB. Codons identified as under positive selection more than one ML methods. Positively selected sites detected by Datamonkey are represented in underlined letters

^bAmino acid substitution occurred along clades, with detailed information marked in Fig. 1

^cRadical changes in amino acid properties under category 6–8 were detected in TreeSAAP. Amino acid sites that belong to the Type II class (greater than 6 property changes) are represented in bold. Physicochemical amino acid properties available in TreeSAAP are as following: α_c : Power to be - C-term., α -helix; α_n : Power to be in the N-terminal of an α -helix; B_r : Buriedness; c : Composition; C_{α} : Helical contact energy; E_t : Long-range non-bonded energy; E_{sm} : Short and medium range non-bonded energy; E_t : Total non-bonding Energy; F : Mean r.m.s. fluctuation displacement; h : Hydropathy; H_{nc} : Normal consensus hydrophobicity; H_{pr} : Surrounding hydrophobicity; H_t : Thermodynamic transfer hydrophobicity; K° : Compressibility; μ : Refractive index; M_w : Molecular weight; M_{wr} : Molecular weight; N_s : Average number of surrounding residues; P_{α} : α -helical tendencies; P_{β} : β -structure tendencies; P_c : Coil tendencies; P : Turn tendencies; p : Polarity; pH_i : Isoelectric point; $\rho K'$: Equilibrium Constant of ionization for COOH; P_r : Polar requirement; R_{ar} : Solvent accessible reduction ratio; R_i : Chromatographic index; V° : Partial specific volume

of high energetic cost for electric organ discharge generation. Although this branch was only detected in the Mb gene with three positively selected codons (Mb: N98A, G12P; Hbb: Q133C), three or more radical property changes occurred at each amino acid (see Table 4). That is to say, globin genes may have adaptively enhanced oxygen binding and transportation in accordance with the changes of high energetic cost during the early evolutionary phase of EOs in gymnotiformes.

Functional consequences of amino acid replacement

Although gymnotiform globin genes contain putatively positively selected sites, it is important to assess their functional relevance. We thus analyzed selected residues for their structural properties to predict potential functional implications. We found that all 26 radical amino acid changes residues were concordant between three methods and thus constitute robust candidates for positive selection (see Table 4). Radical substitutions of amino acids at key positions may change the properties of the molecule. For example, residue 83 in Hba is located very close to the proximal histidine, which is implicated in the iron-proximal histidine linkage [51]. Consequently, substitution at this site seems to be essential in the maintenance of heme oxygen binding.

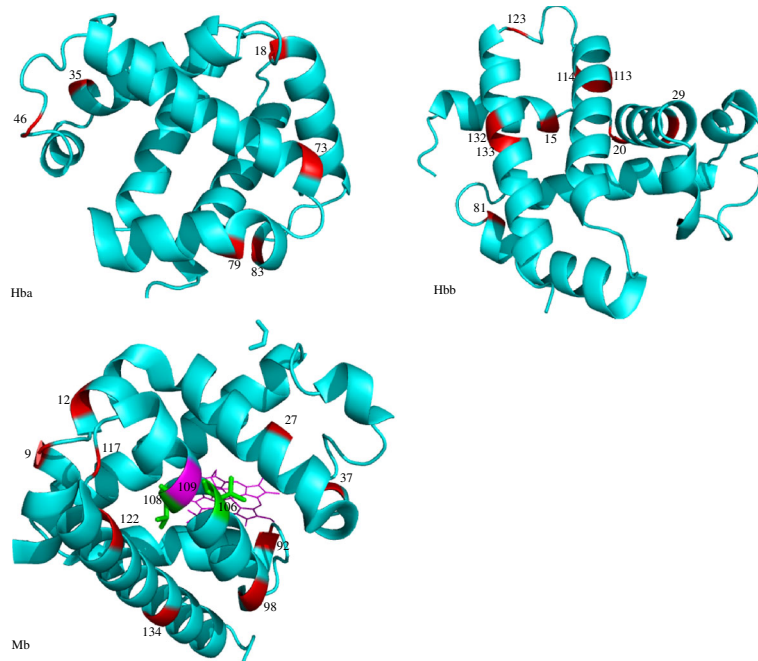


Fig. 2 Radical changes of sites under selection are shown in crystal structure of a: Hba, b: Hbb, and c: Mb with red. The site in Mb identified along lineages leading to *Brachyhypopomus gauderio* and *Electrophorus electricus* is marked with pink (residue 109). Cys substitutions of *Electrophorus electricus* at Mb gene are marked with green (residue 106 and 108)

Furthermore, position 144 in Hbb is close to 147 His. It has been reported that His-HC3(147) β carp (*Cyprinus carpio*) hemoglobin plays a key role in the Root effect, which is a phenomenon associated with non-cooperative oxygen binding and decreased oxygen affinity [52]. Hence, we suggest that positive selection acting on this site is likely to be involved in modulating hemoglobin oxygen combination and cooperation.

For Mb, site 134 is mainly responsible for formation of hydrogen bonds with water in a hydrophobic environment [53]; therefore, amino acid changes at this site are likely to be involved in the regulation of water bonds. In spite of the evidence for selection documented here, functional studies of these candidate positively selected sites are necessary in gymnotiforms in the future.

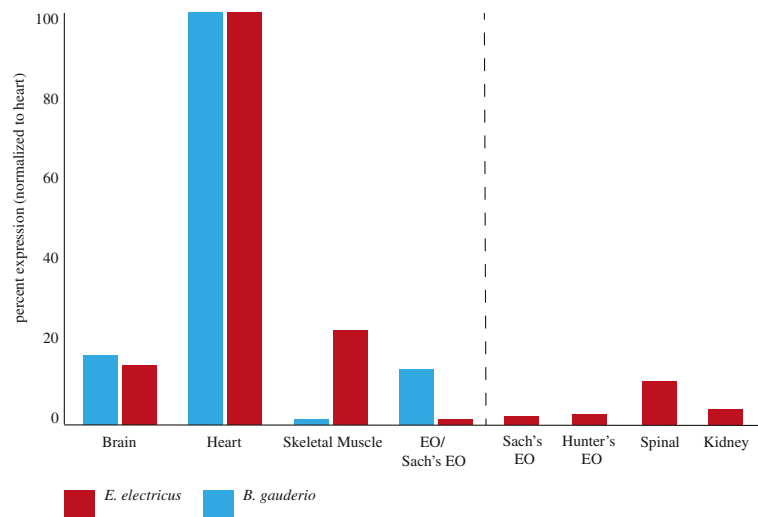
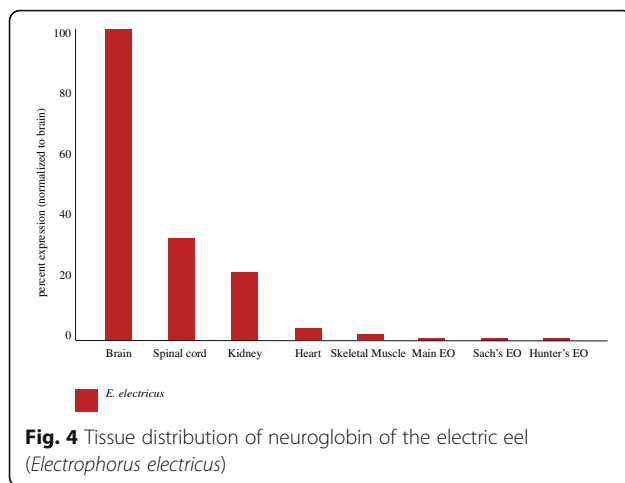


Fig. 3 Tissue distribution of myoglobin of the electric eel (*Electrophorus electricus*) and the pintail knifefish (*Brachyhypopomus gauderio*)



Myoglobin and NO production

Mb plays a pivotal role in the response to hypoxia. On the one hand, Mb facilitates O₂ diffusion from the blood into skeletal and heart muscle of vertebrates; on the other hand, deoxy-Mb may act as a nitrite reductase producing NO from NO₂ in response to cellular hypoxia [13]. Recent studies reported that cysteine (Cys) S-nitrosation of trout and salmon Mb increases heme O₂ affinity, and this allosteric effect may promote hypoxia-induced NO delivery in the heart and improve myocardial efficiency [54, 55]. Interestingly, *Electrophorus electricus* also contains four reactive cysteine (Cys) residues (see Fig. 5). Cys 10 and 108 are identical to trout and salmon, and Cys 106 and 131 are radical amino acid changes from ancestral Val to Cys (V106C, V131C). Moreover, Cys 106 is also a species-specific site in *E. electricus*, and Cys 108 is identified as a positively selected site. It is worth noting that S-nitrosation at reactive cysteines is functionally indispensable for generating S-nitroso Mb (Mb-SNO) and contributing further to NO homeostasis. Hence, it is reasonable to deduce that reactive Cys sites in *E. electricus* may enhance Mb function in oxygen storage and NO delivery during hypoxia.

Interestingly, the Mb of *Gymnarchus niloticus* a member of the other independently evolved group of electric fishes, the mormyroidea, shares Cys 106 (see Fig. 5, blue) with *E. electricus* at a site that is conserved among other teleosts. Whereas all other (200+ species) of mormyroid fishes utilize gills and cannot breathe air under hypoxic conditions [56–58], *G. niloticus* is an air-breathing fish with a highly vascularized gas bladder. These fish also grow to be large (>1.5 meters) similar to *E. electricus*. Therefore, it appears that the shared Cys 106 is a functional convergence in two large, air-breathing electric fish.

Relationship of globin evolution to gymnotiform life histories

The two species with strongest signature of positive selection of globins—*E. electricus* and *B. gauderio*—live

comfortably in hypoxic and even anoxic environments. *E. electricus*, which is the largest gymnotiform capable of reaching ~2 m in length, is unusual among gymnotiforms as it is an obligate air-breather. It surfaces every few minutes to gulp air, which it stores in its mouth and, if prevented from breathing air, it will drown [59]. *E. electricus* is unique even among air breathing fishes as it obtains oxygen from elaborated papillae in the mouth not related to the gills; indeed, its gills are small and underdeveloped. Furthermore, its circulatory system is unlike that of other teleosts in that the oxygenated blood from the oral papillae mixes with the venous circulation before being pumped out of the heart resulting in poorly oxygenated blood. The hematocrit, Hb content, and oxygen capacity of the blood, as well as heart rate and volume of blood moved per unit time are all higher than in most teleosts and may be adaptations to overcome the poor oxygenation of mixed arterial-venous blood [59].

Fish of the genus *Brachyhypopomus* inhabit hypoxic/anoxic waters [10] and in the laboratory, tolerate >6 h of anoxia. They have large well-developed gills, but are also facultative air-breathers either gulping bubbles of air at the surface then descending, or “skulking” at the surface with open mouths taking in air [8]. These two behaviors differ from ASR in that they involve taking in air, whereas ASR is merely taking in well-oxygenated water from the surface layer. Fish that hold air in their mouths must have gills that do not collapse or they would provide too little useful surface area for oxygen absorption.

Gymnotus species are also capable of tolerating >6 h of anoxia by gulping air and storing it in a highly vascularized, extended, lung-like posterior swim bladder unique among gymnotiforms [60]. They use this swim bladder “lung” to extend their aerobic range [61]. The swim bladder is contacted by branches of the celiac artery and hepatic portal systems as well as segmental arteries and veins that then penetrate the body musculature. Although we tried multiple primers to amplify *Gymnotus omarorum* Mb from muscle, heart and EO tissue, we failed to amplify this gene from RNA samples even though we obtained the Hba and Hbb genes from the same samples. Recently, Macqueen et al. [62] revealed that a cardiac Mb deficit has evolved repeatedly in teleosts under diverse ecological settings; in sticklebacks (*Gasterosteus aculeatus*) Mb is even a pseudogene. Although we have not examined *Gymnotus* DNA for a possible pseudogene, we suspect that the Mb gene was lost in *G. omarorum*. In keeping with this, we did not note strong selection on Hbs in this genus. We suggest that their blood is adequately oxygenated in their unique vascularized swim bladder with no further selection pressure to alter Hb affinity for oxygen.

Crampton [10] included *E. virens* among the species that are capable of inhabiting hypoxic environments. However, this species is just barely able to survive in

	10	
Danio rerio	MADHDLVLCGCAVEADYVAANGCEVLRLEFPEYDPLKLFPPKFSISQGLAGSFAVAAH	60
Salmo salar	MANYDMVLCGCPVEADYNNHGGVLSRLEFAEHPETLTLFPKFGAIAAGLSGNAVAAH	60
Thunnus albacares	MADFDVLCGCPVEADYTMGGVLRLEFHEPDTQKLFPPKFGAIAQADLAGNAISAH	60
Oncorhynchus mykiss	MADFDVLCGCPVEADYNNHGGVLSRLEFAEHPETLTLFPKFGAIAAGLSGNAVAAH	60
clade h	MSDFDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGNLAGNADIAAH	60
Electrophorus electricus	MSIDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGNLAGNADIAAH	60
Brachyopomus gauderio	MSDFDMLKCPVEANLKDYGGEVLRLEFHEPDTQKLFPPKFSIPRGNLAGNADIAAH	60
Eigenmannia virescens	MADFDLILKCPVEADYNYGGEVLRLEFHEPDTQKLFPPKFSIPRGSVGNPDIAAH	60
Sternopygus macrurus	MADFDLILKCPTEADYNYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Apteronotus albifrons	MADFDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGNLAGNADIAAH	60
Gymnorhamphichthys sp.	-----VRLFEHPDTQKLFPPKFSIPRGLAGNADIAAH	60
Microsternarchus bilineatus	-----LGRFLDHPDQKFPFQKTRAMIPGNADIAH	34
Rhamphichthys marmoratus	-----VFPVEADTFGGEVLRLEFHEPDTQKLFPPKFSIPRGNLAGNADIAAH	51
Steatogenys elegans	-----VPAVEADTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	51
Parapteronotus hasemani	MADFDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGNLAGNADIAAH	60
Brienomyrus brachyistius	MSDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Campylomormyrus compressirostris	MSDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Gnathonemus petersii	MADFDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Paramormyrops gabonensis	MSDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Petrocephalus soudanensis	MSDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Gymnarchus niloticus	MADYDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Ictalurus punctatus	MSDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Malapterurus electricus	MTDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Bunocephalus coracoides	MADFDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Clarias batrachus	MSDFDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Astyanax mexicanus	MADFDVLCGCAVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Chanos chanos	MADFDVLCGCAVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Carassius auratus	MADYERFLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Clupea harengus	MADFDLILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Cyprinus carpio	MADYERFLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Chitala chitala	MADYDVLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Pantodon buchholzi	MADYDVLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Scleropages formosus	MADHDSVLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Trichogaster trichopterus	MSDYDILKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Amatitlania nigrofasciata	MADYDVLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Anguilla anguilla	--FVLDLAKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	57
Oreochromis niloticus	MADFDVLCGCAVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Takifugu rubripes	MADFDVLCGCAVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
Tetraodon nigroviridis	MGDFDMLKCPVEADYTYGGEVLRLEFHEPDTQKLFPPKFSIPRGLAGNADIAAH	60
	106 108	
Danio rerio	GATVLRKLGELLKAGDHAAALPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Salmo salar	GATVLRKLGELLNARGDHAATLKLPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Thunnus albacares	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Oncorhynchus mykiss	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
clade h	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	119
Electrophorus electricus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Brachyopomus gauderio	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Eigenmannia virescens	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Sternopygus macrurus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Apteronotus albifrons	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Gymnorhamphichthys sp.	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Microsternarchus bilineatus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	94
Rhamphichthys marmoratus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	111
Steatogenys elegans	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	111
Parapteronotus hasemani	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Brienomyrus brachyistius	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Campylomormyrus compressirostris	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Gnathonemus petersii	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Paramormyrops gabonensis	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Petrocephalus soudanensis	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Gymnarchus niloticus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Ictalurus punctatus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Malapterurus electricus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Bunocephalus coracoides	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Clarias batrachus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Astyanax mexicanus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Chanos chanos	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Carassius auratus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Clupea harengus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Cyprinus carpio	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Chitala chitala	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Pantodon buchholzi	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Scleropages formosus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Trichogaster trichopterus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Amatitlania nigrofasciata	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Anguilla anguilla	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Oreochromis niloticus	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Takifugu rubripes	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
Tetraodon nigroviridis	GATVLRKLGELLKAGDHAAILKPLANTHANHVKVANNFNRLITETLIVKVAEKAGLDA	120
	131	
Danio rerio	QGQALRRVMDVIGDIDGYKELGFG	147
Salmo salar	QGEALRRVMDVIGDIDGYKELGFG	147
Thunnus albacares	QGTALRRVMDVIGDIDGYKELGFG	147
Oncorhynchus mykiss	QGEALRRVMDVIGDIDGYKELGFG	146
clade h	QGTALRRVMDVIGDIDGYKELGFG	147
Electrophorus electricus	QGTALRRVMDVIGDIDGYKELGFG	147
Brachyopomus gauderio	TAEAFRRVMDVIGDIDGYKELGFG	147
Eigenmannia virescens	TQDALRRVMDVIGDIDGYKELGFG	147
Sternopygus macrurus	TQDALRRVMDVIGDIDGYKELGFG	147
Apteronotus albifrons	TQDAVRRVMDVIGDIDGYKELGFG	145
Gymnorhamphichthys sp.	TQDALRRVMDVIGDIDGYKELGFG	112
Microsternarchus bilineatus	TQDALRRVMDVIGDIDGYKELGFG	113
Rhamphichthys marmoratus	TQDALRRVMDVIGDIDGYKELGFG	129
Steatogenys elegans	TQDALRRVMDVIGDIDGYKELGFG	129
Parapteronotus hasemani	TQDAVRRVMDVIGDIDGYKELGFG	147
Brienomyrus brachyistius	QGTALRRVMDVIGDIDGYKELGFG	147
Campylomormyrus compressirostris	QGTALRRVMDVIGDIDGYKELGFG	147
Gnathonemus petersii	QGTALRRVMDVIGDIDGYKELGFG	145
Paramormyrops gabonensis	QGTALRRVMDVIGDIDGYKELGFG	147
Petrocephalus soudanensis	QGQALRRVMDVIGDIDGYKELGFG	147
Gymnarchus niloticus	TQTAMRNILAVIGDIDGYKELGFG	147
Ictalurus punctatus	QGDAMRRVMDVIGDIDGYKELGFG	147
Malapterurus electricus	QGDALRRVMDVIGDIDGYKELGFG	147
Bunocephalus coracoides	TQDALRRVMDVIGDIDGYKELGFG	145
Clarias batrachus	QGDALRRVMDVIGDIDGYKELGFG	147
Astyanax mexicanus	QGDALRRVMDVIGDIDGYKELGFG	147
Chanos chanos	QGDALRRVMDVIGDIDGYKELGFG	147
Carassius auratus	QGDALRRVMDVIGDIDGYKELGFG	147
Clupea harengus	QGDALRRVMDVIGDIDGYKELGFG	147
Cyprinus carpio	QGDALRRVMDVIGDIDGYKELGFG	147
Chitala chitala	QGDALRRVMDVIGDIDGYKELGFG	145
Pantodon buchholzi	QGDALRRVMDVIGDIDGYKELGFG	145
Scleropages formosus	QGDALRRVMDVIGDIDGYKELGFG	145
Trichogaster trichopterus	QGDALRRVMDVIGDIDGYKELGFG	144
Amatitlania nigrofasciata	QGDALRRVMDVIGDIDGYKELGFG	144
Anguilla anguilla	QGDALRRVMDVIGDIDGYKELGFG	139
Oreochromis niloticus	QGDALRRVMDVIGDIDGYKELGFG	147
Takifugu rubripes	QGDALRRVMDVIGDIDGYKELGFG	147
Tetraodon nigroviridis	QGDALRRVMDVIGDIDGYKELGFG	147

Gymnotiformes

Mormyroidea

Gymnotiformes

Mormyroidea

Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Multiple sequences alignment of 40 teleost Mb gene, showing Cys residues (i.e., 10, 106, 108, and 131) of *Electrophorus electricus* (pink). The common substitutions identified are located at the amino acid position of the zebrafish (*Danio rerio*) ortholog. The consensus numbering of zebrafish is given in the right. The Cys 10 and 108 residues of *Salmo salar* and *Thunnus albacares* are shadowed by purple. The corresponding Cys sites in *Gymnarchus niloticus* (Mormyroidea) are shown in blue, in Gymnotiformes in green, in the remaining species in orange. The alignments of each Cys sites are in the red boxes. Clade h (Fig. 1) is the common ancestor of the genus *Gymnotus* and *Electrophorus electricus*. NCBI accession numbers are provided in Additional file 1: Table S1

hypoxia. They tend to avoid hypoxic water although they are capable of surviving an hour or so of anoxic conditions using ASR [10], but they do not gulp air. They are sensitive to hypoxia under which the amplitude of their EODs decreases rapidly [11]. In accordance with this, we observe only minimal evidence of positive selection on Hbb in this species.

The other species in this study, which show little or only weak evidence of positive selection on globin genes, do poorly in hypoxic conditions and cannot survive anoxia [10]. Under anoxic conditions in the laboratory they respond with ASR as long as they are able and eventually fall immobile to the bottom. These include all the wave species and some of the pulse-type fish.

Finally, the genera *Brachyhypopomus* and *Gymnotus* are speciose [4] with member species distributed throughout the gymnotiform's whole range. For example, fish from these two genera are the primary species in the most Southern extent of the gymnotiform's distribution in Uruguay [63, 64]. The evolution of different modes of air-breathing may have given them an advantage over those species unable to derive oxygen from the air.

Expression of Mb and Ngb in *E. electricus* and *B. gauderio*

As in some other teleosts, Mb is most highly expressed in heart (although this varies considerably across species [62]) and less so in muscle. However, as Mb is expressed in “slow” oxidative but not “fast” non-oxidative muscle fibers, and slow fibers are small and few in number in teleost muscle, its mRNA will not be abundant in a sample of epaxial muscle even if it is highly expressed in slow muscle [65, 66]. Thus, the difference in muscle Mb expression between *E. electricus* and *B. gauderio* may depend on composition of the muscle sub-types and the exact location from which samples were obtained.

Our expectation was that Mb levels would be high in the EO consistent with the ongoing EOD activity. But Mb levels are modest in the EO of both species. We did not note a difference in expression between the Sach's organ and Hunter's/Main EO of *E. electricus*. Sach's EO constantly fires a low voltage-pulse at a low frequency (~1–10 Hz), whereas all three EOs discharge in high frequency bursts of >400 Hz when fish are capturing prey or defending themselves [67]. *B. gauderio* discharges at rest at ~15 Hz and during social interactions or foraging at 30–40 Hz, with

occasional high-frequency bursts (100–200 Hz) called “chirps” that are aggressive signals [68].

There are a few possible and not mutually exclusive explanations for the low levels of Mb in the EO. First, EO originates from fast muscle fibers, and fast muscle fibers do not express Mb; there might be a developmental constraint minimizing Mb expression in the EO. Second, some gymnotiforms possess mechanisms to decrease EOD amplitude (and therefore oxygen consumption) during anoxia [11]. This has only been studied in a few species and it would be intriguing to compare the distribution of EO Mb expression, EOD “usage” (discharge rate), and occurrence of mechanisms for reducing EOD amplitude in the face of anoxia in the EO across gymnotiforms. Perhaps different species trade off EO Mb concentration with extent of anoxia-dependent reduction of the EOD amplitude. Third, EOs are well vascularized and it may not be necessary to have a local reserve source of O₂. Fourth, it is possible that EOs switch to anaerobic metabolism during and after periods of high activity. At this point there is no evidence for or against this. It is also worth noting that Mb expression is high in heart and low in muscle in the elephant shark suggesting that this is the ancestral vertebrate pattern [69].

Recently, it has been recognized that Mb is expressed in the brain and kidney as well as in a few other non-muscle tissues in teleosts. Mainly, Mb is expressed in the endothelial cells of blood vessels in the brain and gills, and the epithelial cells of the tubules in the kidney [65]. The expression of Mb in the brain in *E. electricus* and *B. gauderio* and the kidney of *E. electricus* is in accord with these observations.

Ngb is primarily expressed in neural tissues in mammals and teleosts. It is reported to be in the gills in zebrafish [70]. As there are no gill transcriptomes available of *E. electricus* we do not know if it is expressed in the gills of this species. Ngb was expressed in *E. electricus* kidney but zebrafish kidney has not yet been examined. Finally, it is not expressed in zebrafish muscle nor did we observe it in *E. electricus* muscle. Qualitatively, the expression pattern of Mb and Ngb does not differ from non-gymnotiforms. Thus, it appears that the main effect of selection on *E. electricus* and *B. gauderio* Mb is on its sequence rather than its expression pattern in individuals housed under normoxic conditions. Further experiments would be necessary to determine whether there

are additional adaptations on the regulation of globin genes under different environmental conditions.

Conclusion

The gymnotiforms arose in Gondwana [71] and diverged rapidly into lineages with differences in EO morphology and signaling. Some amino acid substitutions occurred in myoglobin genes as the EO was initially evolving, perhaps as a result of the new energy demands imposed by electrogenesis. While many gymnotiforms rely on gills for oxygen uptake, a few lineages evolved distinct modes of air-breathing. Along with this, we note variation in the strength of selection on different globin genes depending on the mode of air breathing. We anticipate that future studies will elucidate the physiology and biochemistry of these globin genes and seek other molecules under selection in the oxygen transport or metabolic pathways underlying energy usage.

Additional files

Additional file 1: Table S1. List of Species Used in this Study. **Table S2.** Statistics for amplified genes for each species. **Table S3.** List of primers used to amplify the coding regions of globin genes in this study. **Table S4.** Positive Selection Detected by Site Models in PAML. **Table S5.** Positive Selection Detected by Branch Models in PAML. **Table S6.** Raw and Processed Reads from Each Tissue. **Table S7.** Evaluation of transcriptome completeness against the Vertebrates, Metazoans and Eukaryotes datasets. (DOCX 42 kb)

Additional file 2: Figure S1. Sequences alignment of globin genes. **Figure S2** The phylogenetic trees constructed from nucleotide sequences used maximum likelihood tests. (ZIP 729 kb)

Additional file 3: Further details of *Brachyhypopomus gauderio* transcriptome. (DOCX 133 kb)

Acknowledgements

The authors acknowledge Drs. Christopher Braun (Hunter College, NYC), G. Troy Smith (Indiana University), Ana Silva (Universidad de la República de Uruguay), Michael Markham (University of Oklahoma, Norman), and Hernán Lopez-Fernandez (Royal Ontario Museum) for contributing RNA sequences or tissues for PCR for this study. This work was supported by the China Scholarship Council (RT), the University of Texas (HZ), NSF grant IOS 1557857 (HZ), Nanjing Normal University (GY), NSF grants 31325025 and 31630071 (GY), Nanjing Normal University (RT) and Priority Academic Program Development of Jiangsu Higher Education Institutions (RT).

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its Additional files 1, 2, and 3.

Authors' contributions

HZ, RT, and GY conceived and designed the study. RT and ML collected the data. YL prepared the samples and submitted the sequences. RT performed the laboratory work. RT analyzed the data. All authors contributed to writing the manuscript and approved the final version of it.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

As the research did not include living or anesthetized animals, no ethical approval was necessary. All methods used to collect observational data were

non-invasive, were in compliance with the regulations and guidelines of the University of Texas at Austin's Institutional Animal Care and Use Committee.

Author details

¹Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China. ²Department of Integrative Biology, The University of Texas, Austin, TX 78759, USA. ³Department of Neuroscience, The University of Texas, Austin, TX 78759, USA. ⁴Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA.

Received: 5 July 2016 Accepted: 26 January 2017

Published online: 13 February 2017

References

- Von der Emde G. Electroreception. In: Bullock TH, Fay CDH, Richard R, editors. Neurosciences-from molecule to behavior: a university textbook. Springer: Springer Science & Business Media; 2013. p. 409–25.
- Caputi AA, Carlson BA, Macadar O. Electric organs and their control. In: Electroreception. New York: Springer; 2005. p. 410–51.
- Hopkins CD. Neuroethology of electric communication. *Ann Rev Neurosci*. 1988;11:497–535.
- Tagliacollo VA, Bernt MJ, Craig JM, Oliveira C, Albert JS. Model-based total evidence phylogeny of Neotropical electric knife-fishes (Teleostei, Gymnotiformes). *Mol Phylogenet Evol*. 2016;95:20–33.
- Salazar VL, Krahe R, Lewis JE. The energetics of electric organ discharge generation in gymnotiform weakly electric fish. *J Exp Biol*. 2016;216:2459–68.
- Julian D, Crampton WGR, Wohlgemuth SE, Albert JS. Oxygen consumption in weakly electric Neotropical fishes. *Oecologia*. 2003;137:502–11.
- Crampton WGR. Gymnotiform fish: an important component of Amazonian floodplain fish communities. *J Fish Biol*. 1996;48:298–301.
- Crampton WGR, Chapman LJ, Bell J. Interspecific variation in gill size is correlated to ambient dissolved oxygen in the Amazonian electric fish *Brachyhypopomus* (Gymnotiformes: Hypopomidae). *Environ Biol Fish*. 2008;83:223–35.
- Val AL, Silva MNP, Almeida-Val VMF. Hypoxia adaptation in fish of the Amazon: a never-ending task. *S Afr J Zool*. 1998;33:107–14.
- Crampton WGR. Effects of anoxia on the distribution, respiratory strategies and electric signal diversity of gymnotiform fishes. *J Fish Biol*. 1998;53:307–30.
- Reardon EE, Parisi A, Krahe R, Chapman LJ. Energetic constraints on electric signalling in wave-type weakly electric fishes. *J Exp Biol*. 2011;214:4141–50.
- Sinnett PM, Markham MR. Food deprivation reduces and leptin increases the amplitude of an active sensory and communication signal in a weakly electric fish. *Horm Behav*. 2015;71:31–40.
- Burmester T, Hankeln T. Function and evolution of vertebrate globins. *Acta Physiol*. 2014;211:501–14.
- Hardison RC. A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc Natl Acad Sci U S A*. 1996;93:5675–9.
- Bashford D, Chothia C, Lesk AM. Determinants of a protein fold: Unique features of the globin amino acid sequences. *J Mol Biol*. 1987;196:199–216.
- Rahbar S. Hemoglobin: Structure, function, evolution, and pathology. *Am J Hum Genet*. 1983;35:781.
- Wittenberg BA, Wittenberg JB. Transport of oxygen in muscle. *Annu Rev Physiol*. 1989;51:857–78.
- Dickerson RE, Geis I. Hemoglobin: structure, function, evolution and pathology. Menlo Park: Benjamin-Cummings Publishing Company; 1983.
- Gros G, Wittenberg BA, Jue T. Myoglobin's old and new clothes: from molecular structure to function in living cells. *J Exp Biol*. 2010;213:2713–25.
- Hankeln T, Wystub S, Laufs T, Schmidt M, Gerlach F, Saaler-Reinhardt S, Reuss S, Burmester T. The cellular and subcellular localization of neuroglobin and cytoglobin clue to their function? *IUBMB life*. 2004;56:671–9.
- Vinogradov SN, Moens L. Diversity of globin function: enzymatic, transport, storage, and sensing. *J Biol Chem*. 2008;283:8773–7.
- Offerman JD, Rychlik W. Oligo primer analysis software. In: Krawetz SA, Womble DD (Eds). *Introduction to Bioinformatics. A Theoretical and Practical Approach*. New Jersey: Humana Press Inc.; 2003. p. 345–61.
- Griffin AM, Griffin HG, Olson SA. MacVector: an integrated sequence analysis program for the Macintosh. *Computer Analysis of Sequence Data: Part II*. 1994. p. 195–201.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.

25. Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 2008; 25:1253–6.
26. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307–21.
27. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
28. Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol.* 2003;20:18–20.
29. Yang Z, Wong WSW, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005;22:1107–18.
30. Pond SLK, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21:2531–3.
31. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005;22:2472–9.
32. Pond SLK, Murrell B, Fourment M, Frost SD, Delpoit W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol.* 2011;28(11):3033–43.
33. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics.* 2003;19:671–2.
34. DeLano WL. The PyMOL molecular graphics system. San Carlos: DeLano Scientific; 2002.
35. Bolger AM, Marc L, Bjoern U. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
36. MacManes MD. On the optimal trimming of high-throughput mRNA sequence data. *Front Genet.* 2014;5:13.
37. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol.* 2011;29:644.
38. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013;8:1494–512.
39. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2.
40. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC bioinformatics.* 2009;10:1.
41. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39(Web Server issue):W29–37.
42. Rice P, Ian L, Alan B. EMBOS: the European molecular biology open software suite. *Trends Genet.* 2000;16:276–7.
43. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics.* 2011;12:1.
44. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34:525–7.
45. Gallant JR, Traeger LL, Volkening JD, Moffett H, Chen PH, et al. Genomic basis for the convergent evolution of electric organs. *Science.* 2014;344:1522–5.
46. Traeger LL, Volkening JD, Moffett H, Gallant JR, Chen PH, et al. Unique patterns of transcript and miRNA expression in the South American strong voltage electric eel (*Electrophorus electricus*). *BMC Genomics.* 2015;16:243.
47. Qiu Q, Zhang G, Ma T, Qian W, Wang J, et al. The yak genome and adaptation to life at high altitude. *Nat Genet.* 2012;44:946–9.
48. Projecto-Garcia J, Natarajan C, Moriyama H, Weber RE, Fago A, et al. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc Natl Acad Sci U S A.* 2013;110:20669–74.
49. Tian R, Wang Z, Niu X, Zhou K, Xu S, Yang G. Evolutionary genetics of hypoxia tolerance in cetaceans during diving. *Genome Biol Evol.* 2016;8:827–39.
50. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* 2010;329:75–8.
51. Parkhurst LJ, Goss DJ, Perutz MF. Kinetic and equilibrium studies on the role of the beta-147 histidine in the Root effect and cooperativity in carp hemoglobin. *Biochemistry.* 1983;22:5401–9.
52. Friedman JM, Scott TW, Stepnoski RA, et al. The iron-proximal histidine linkage and protein control of oxygen binding in hemoglobin. A transient Raman study. *J Biol Chem.* 1983;258:10564–72.
53. Teeter MM. Myoglobin cavities provide interior ligand pathway. *Protein Sci.* 2004;13:313–8.
54. Helbo S, Fago A. Allosteric modulation by S-nitrosation in the low-O₂ affinity myoglobin from rainbow trout. *Am J Physiol-Reg I.* 2011;300:R101–8.
55. Helbo S, Gow AJ, Jamil A, Howes BD, Smulevich G, Fago A. Oxygen-linked S-nitrosation in fish myoglobins: A cysteine-specific tertiary allosteric effect. *PLoS One.* 2014;9, e97012.
56. Chapman LJ, Chapman CA. Hypoxia tolerance of the mormyrid *Petrocephalus catostoma*: implications for persistence in swamp refugia. *Copeia.* 1998;998:762–8.
57. Chapman LJ, Hulen KG. Implications of hypoxia for the brain size and gill morphometry of mormyrid fishes. *J Zool.* 2001;254:461–72.
58. Berenbrink M, Koldkjaer P, Kepp O, Cossins AR. Evolution of oxygen secretion in fishes and the emergence of a complex physiological system. *Science.* 2005;307:1752–7.
59. Johansen K, Lenfant C, Schmidt-Nielsen K, Petersen JA. Gas exchange and the control of breathing in the electric eel, *Electrophorus electricus*. *Z Vergl Physiologie.* 1968;61:137–63.
60. Liem KF, Eclancher B, Fink WL. Aerial Respiration in the Banded Knife Fish *Gymnotus Carapo* (teleostei: Gymnotoidei). *Physiol Zool.* 1984;57:185–95.
61. McKenzie DJ, Steffensen JF, Taylor EW, Abe AS. The contribution of air breathing to aerobic scope and exercise performance in the banded knifefish *Gymnotus carapo* L. *J Exp Biol.* 2012;215:1323–30.
62. Macqueen DJ, Johnston IA. Cardiac myoglobin deficit has evolved repeatedly in teleost fishes. *Biol Lett.* 2014;10:20140225.
63. Richer-de-Forges MM, Crampton WGR, Albert JS. A new species of *Gymnotus* (Gymnotiformes, Gymnotidae) from Uruguay: description of a model species in neurophysiological research. *Copeia.* 2009;3:538–44.
64. Giora J, Malabarba LR. *Brachyhypopomus gauderio*, new species, a new example of underestimated species diversity of electric fishes in the southern South America (Gymnotiformes, Hypopomidae). *Zootaxa.* 2009;2093:60–8.
65. Cossins AR, Williams DR, Foulkes NS, Berenbrink M, Kipar A. Diverse cell-specific expression of myoglobin isoforms in brain, kidney, gill and liver of the hypoxia-tolerant carp and zebrafish. *J Fish Biol.* 2009;212:627–38.
66. Jaspers RT, Testerink J, Della Gaspera B, Chanoine C, Bagowski CP, van der Laarse WJ. Increased oxidative metabolism and myoglobin expression in zebrafish muscle during chronic hypoxia. *Biol Open.* 2014;3(8):718–27.
67. Catania K. The shocking predatory strike of the electric eel. *Science.* 2014; 346:1231–4.
68. Silva AC, Perrone R, Zubizarreta L, Batista G, Stoddard PK. Neuromodulation of the agonistic behavior in two species of weakly electric fish that display different types of aggression. *J Exp Biol.* 2013;216:2412–20.
69. Opazo JC, Lee AP, Hoffmann FG, Toloza-Villalobos J, Burmester T, Venkatesh B, Storz JF. Ancient duplications and expression divergence in the globin gene superfamily of vertebrates: insights from the elephant shark genome and transcriptome. *Mol Biol Evol.* 2015;32:1684–94.
70. Fuchs C, Heib V, Kiger L, Haberkamp M, Roesner A, Schmidt M, Hamdane D, Marden MC, Hankeln T, Burmester T. Zebrafish reveals different and conserved features of vertebrate neuroglobin gene structure, expression pattern, and ligand binding. *J Biol Chem.* 2004;279:24116–22.
71. Lavoué S, Miya M, Arnegard ME, Sullivan JP, Hopkins CD, Nishida M. Comparable ages for the independent origins of electrogenesis in African and South American weakly electric fishes. *PLoS One.* 2012;7, e36287.
72. Silva A, Quintana L, Galeano M, Errandonea P. Biogeography and breeding in *Gymnotiformes* from Uruguay. *Environ Biol Fish.* 2003;66:329–38.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

