Report

Current Biology

Modulation of bioelectric cues in the evolution of flying fishes

Highlights

- Aerial gliding evolved in fishes through differential allometric growth of paired fins
- The amino acid transporter *lat4a* and the potassium channel *kcnh2a* regulate fin size
- Convergence and selection in *lat4a* and bioelectric signaling genes in flying fishes
- Interaction between *lat4a* and *kcnh2a* establishes flying fish bauplan in zebrafish

Authors

Jacob M. Daane, Nicola Blum, Jennifer Lanni, ..., Stephen L. Johnson, Nathan R. Lovejoy, Matthew P. Harris

Correspondence

j.daane@northeastern.edu (J.M.D.), harris@genetics.med.harvard.edu (M.P.H.)

In brief

Daane et al. intersect patterns of genome evolution in flying fishes with genetic screens in zebrafish to identify a role for the amino acid transporter *lat4a* and the potassium channel *kcnh2a* in regulating fin size. A combination of alleles from both genes is sufficient to reproduce the flying fish fin bauplan in the distantly related zebrafish.

Daane et al., 2021, Current Biology *31*, 1–10
November 22, 2021
https://doi.org/10.1016/j.cub.2021.08.054

CellPress

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

Current Biology

CellPress

Report

Modulation of bioelectric cues in the evolution of flying fishes

Jacob M. Daane,[1,2,3,9,*] Nicola Blum,[1,2] Jennifer Lanni,[4] Helena Boldt,[1,2] M. Kathryn Iovine,[5] Charles W. Higdon,[6] Stephen L. Johnson,[6,8] Nathan R. Lovejoy,[7] and Matthew P. Harris[1,2,10,11,*]

[1]Department of Orthopaedic Research, Boston Children's Hospital, Boston, MA 02124, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA 02124, USA
[3]Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, Nahant, MA 01908, USA
[4]Department of Biology, Wheaton College, Norton, MA 02766, USA
[5]Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015, USA
[6]Department of Genetics, Washington University Medical School, St. Louis, MO 63110, USA
[7]Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON M1C-1A4, Canada
[8]Posthumous work
[9]Twitter: @jakedaane
[10]Twitter: @fishyskeleton
[11]Lead contact
*Correspondence: j.daane@northeastern.edu (J.M.D.), harris@genetics.med.harvard.edu (M.P.H.)
https://doi.org/10.1016/j.cub.2021.08.054

SUMMARY

Changes to allometry, or the relative proportions of organs and tissues within organisms, is a common means for adaptive character change in evolution. However, little is understood about how relative size is specified during development and shaped during evolution. Here, through a phylogenomic analysis of genome-wide variation in 35 species of flying fishes and relatives, we identify genetic signatures in both coding and regulatory regions underlying the convergent evolution of increased paired fin size and aerial gliding behaviors. To refine our analysis, we intersected convergent phylogenomic signatures with mutants with altered fin size identified in distantly related zebrafish. Through these paired approaches, we identify a surprising role for an L-type amino acid transporter, *lat4a*, and the potassium channel, *kcnh2a*, in the regulation of fin proportion. We show that interaction between these genetic loci in zebrafish closely phenocopies the observed fin proportions of flying fishes. The congruence of experimental and phylogenomic findings point to conserved, non-canonical signaling integrating bioelectric cues and amino acid transport in the establishment of relative size in development and evolution.

RESULTS AND DISCUSSION

The evolution of fin allometry in Beloniformes

Several fish lineages have independently evolved elongated, wing-like fins that enable aerial gliding.[1] The most accomplished aerial gliders are members of the family Exocoetidae (Beloniformes), or the "flying fishes." All flying fish species have enlarged pectoral fins (Figure 1A) that they use as airfoils to glide through the air to avoid the aquatic predators of open-water epipelagic habitats.[1] Flying fishes also have a hypocercal caudal fin with a stiffened and elongated ventral lobe (Figure 1A). The hypocercal caudal fin facilitates the initial emergence and take-off from the water, as well as allowing some species to generate additional momentum without fully re-entering the water.[2] In addition to flying fishes, several partial-gliding species occur within the closely related halfbeak family (Hemiramphidae; Figures 1B and 2C). "Flying halfbeaks" of the genera *Oxyporhamphus* and *Euleptorhamphus* have elongated pectoral fins and exhibit gliding behaviors, though these fishes fail to attain the degree of controlled, prolonged flight associated with Exocoetidae. Species of the flying halfbeak genus *Oxyporhamphus* have also lost the halfbeak jaw morphology, similar to flying fishes.[2–4] The evolution of fin proportion within this clade presents a unique window to address the changes associated with adaptive locomotion and allometry.

To understand the changes in fin allometry during beloniform evolution, we measured fin length as a function of body size (standard length) in representative species across the clade (Figure 1B; Data S1A). While halfbeaks and needlefishes show subtle increases in pectoral fin length as the body grows, the length of flying fish pectoral fins increases rapidly (Figure 1B; Data S1A). The allometric relationship between body size and pectoral fin length was identical across the two- and four-wing flying fishes (Figure 1B), indicating that the scaling of fin proportion may be fixed across all flying fishes. Intriguingly, the two genera of gliding halfbeaks (*Euleptorhamphus* and *Oxyporhamphus*), which exhibit an increase in fin size that is convergent with flying fishes, show a fin-body length allometric scaling relationship that is intermediate between halfbeaks and flying fishes (Figures 1B and 1C; Data S1A).[1]

Identification of genetic variation across Beloniformes

To look for genetic trends shared among flying fishes and gliding halfbeaks, we used a multi-species, cross-species targeted

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

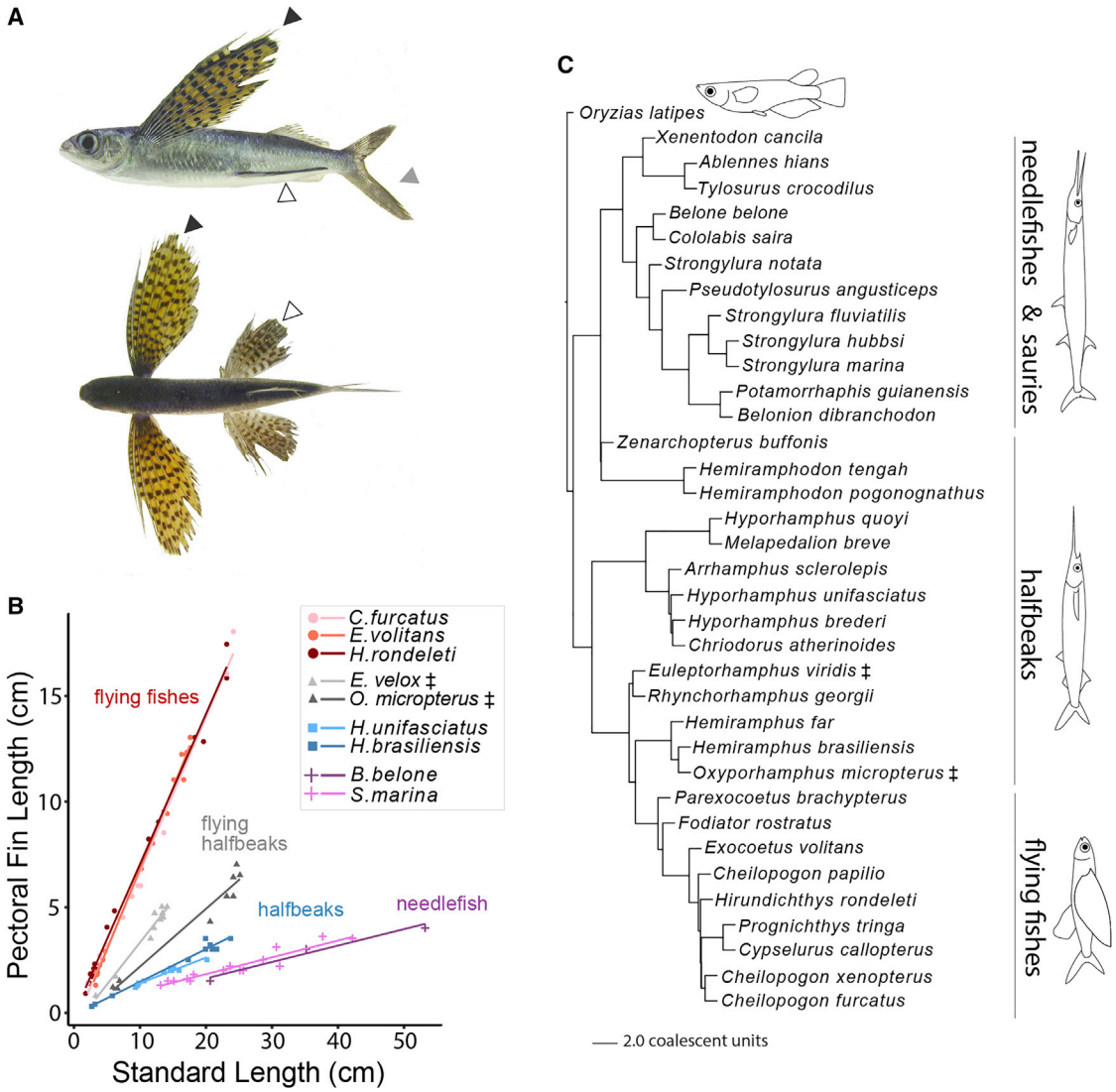

Figure 1. Evolution of fin allometry and gliding behavior in Beloniformes

(A) Lateral and dorsal views of a flying fish, *Cypselurus callopterus*, highlighting elongated pectoral (black arrowhead) and pelvic fins (white arrowhead) that act as an airfoil to enable aerial gliding behavior. The asymmetry of the caudal fin (hypocercal) aids in above-water propulsion (gray arrowhead).

(B) Static allometry of pectoral fins in flying fishes (Exocoetidae; *C. furcatus*, *E. volitans*, and *H. rondeleti*) compared to sauries and needlefish (*B. belone* and *S. marina*) and halfbeaks (*H. unifasciatus* and *H. brasiliensis*). The "flying halfbeaks" (*E. velox* and *O. micropterus*) also exhibit aerial gliding behavior and have an intermediate pectoral fin length relative to the fishes of Exocoetidae.

(C) Phylogeny of the beloniforms sequenced in this study. ‡ indicates lineages of flying halfbeak. All nodes were supported with a quadpartition posterior probability of 1.00.

sequencing design to sample over 300,000 distinct genetic loci representing conserved non-coding elements (CNEs) and protein-coding genes. We extracted and pooled genomic DNA from several individuals from each of 35 different beloniform species, including two- and four-winged flying fishes, halfbeaks, sauries, and needlefishes (Table S1; Data S1B). We obtained coverage from an average of 78.3% of the targeted genomic regions at an average 38.3-fold depth (Tables S2 and S3). This recovery was impressive, given, and despite, the large evolutionary distance between the species sequenced and the reference genomes used to design targeted sequencing baits (>70 Ma).[5–7] We found comparable recovery among CNE, microRNA, and coding sequence categories (Table S2).

As with other cross-species targeted sequence capture experiments,[8,9] there is enrichment within poorly covered regions for genes that are rapidly evolving, particularly in gene clusters involving the immune system (Data S1C). Though these gene classes are enriched for missing coverage compared to the rest of the genome, importantly, most of the genes (77.9%) within these GO terms do not have low coverage (Data S1C).

By sequencing pooled DNA samples from several individuals for each species (Table S1; Data S1B), we were able to

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

categorize fixed and heterozygous nucleotide variation for each species (Data S1D). Additionally, among reconstructed exons, we found no evidence for genome-wide differences in copy number variation between species (Data S1E). With these broad genetic data, we estimated a phylogeny for the clade from 4,683 reconstructed gene trees in our dataset, concordant with published phylogenies (Figure 1D).[2,4,10]

Genomic imprint of aerial gliding in fishes

To detail the genetic pathways associated with gliding behavior and morphology, we analyzed patterns of evolutionary rate across the genomes of gliding compared to non-gliding lineages of Beloniformes (Figure 2A; Data S1F). Randomly sampled controls representing species groupings with the same phylogenetic distributions (tip and ancestral branches) as in gliding lineages do not show evidence for accelerated or constrained evolution across any gene groups (Figure S1B). In gliding branches, however, we observed an elevated evolutionary rate in several gene classes. Many of the functions of these gene classes associate with key morphological adaptations thought to be important for gliding (Figure 2A; Data S1F). Notably, accelerated gene groupings include fin development and morphogenesis, pectoral fin morphogenesis, embryonic digit morphogenesis, and fore- and hindlimb morphogenesis (Figure 2A; Data S1F). Beyond elongated fins, above-water taxiing and aerial gliding are enabled by adaptations that affect balance and musculature. These morphological changes include enlarged semicircular canals,[11] increased size of pectoral fin musculature and neuromuscular adaptations to enable propulsive tail oscillations at speeds up to 50 beats per second,[1,12] and pyramidal corneas adapted to provide clear image resolution both above and below the water surface.[13] Accordingly, we observed an accelerated sequence evolution across a number of gene groupings functionally relating to these traits, including vestibular receptor cell development, semicircular canal formation and morphogenesis, adult locomotory behavior and directional locomotion, hindbrain development and cerebellar granule cell migration, camera-type eye morphogenesis, as well as muscle hypertrophy (Figure 2A; Data S1F). Intriguingly, we find accelerated sequence evolution in the genes involved in locus cereleus development, a key region of the brain involved in alertness, anxiety, and panic response in humans,[14] as well as epinephrine secretion and transport and adrenergic receptor activity (Figure 2A; Data S1F). These results are consistent with the hypothesis of gliding evolution as an anti-predator response.[1] Thus, through this comparative analysis, we find genes associated with key gliding adaptations under accelerated sequence evolution in gliding beloniforms.

Genetic basis of fin allometry

To further explore the genetic regulators of fin allometry in flying fishes, we intersected forward genetic and comparative genomic approaches. We leveraged an existing mutant collection in zebrafish with fin size phenotypes to ask whether loci identified in experimental screens could refine the analysis of genomic variation toward isolating mechanisms affecting natural variation in size. This analysis benefited from the identification of two mutants with changes in fin proportion in genes whose function was previously not known to regulate size and growth.

Identification of leucine transporter lat4a as a regulator of fin size

We identified a dominant mutant, *nr21*, as having small adult fins compared to wild-type siblings (Figures 3A and 4). In addition to overall fin size, the length of lepidotrichia segments in *nr21* are smaller than in wild-type fish (Figures S2A and S2B). The *nr21* mutant, however, does not have a statistically decreased body size (Figure 3B). Using a heterogeneity-based mapping approach,[15] we mapped *nr21* to chromosome 15 and identified a nonsynonymous mutation (T200K) in a conserved domain of the L-type amino acid transporter, *lat4a* (Figures 3C and S2D–S2F). Mammalian Lat4 shuttles isoleucine, leucine, methionine, and phenylalanine into the cell in an ion- and pH-independent manner.[16] Lat4 has been linked to overall body size in mice,[17] but its role in proportion or patterning is not known. To confirm that lat4a causes the *nr21* phenotype, we generated frameshift mutations in *cis* to the *nr21/lat4a* allele though CRISPR-Cas9-mediated gene editing. These induced mutations resulted in reversion of *nr21* to a wild-type fin phenotype, indicating that the mutation in $lat4a^{nr21}$ results from a gain-of-function effect (Figures 3C and 3D); conversely, we find that zebrafish lacking *lat4a* have no fin phenotype (Figures S2G–S2I).

The zebrafish longfin mutant and the regulation of fin proportion

The dominant *longfin* mutant (lof^{dt2}) was first isolated in an aquarium population and is one of the earliest identified zebrafish mutants (*Tüpfel long fin*).[18] The *lof* mutant has coordinated overgrowth of all fins that arise during late development. The effect of the mutant allele on growth is dose sensitive such that homozygous fish have larger fins (Figure 3E). Importantly, *lof* mutants maintain elongate but regular segmented lepidotrichia (Figures S2A and S2B), in contrast to the $kcnk5b^{alf}$ fin overgrowth mutant having irregular segmentation (Figures S2A and S2B).[19] As in the *lof* mutant, flying fish segmentation pattern is regularly spaced (Figure S2C).

We investigated the genetic causes of the *lof* phenotype. Positional mapping showed limited ability to refine the mapping interval of *lof*, in line with previous reports.[20] Using both ENU and γ-ray mutagenesis, we performed two independent reversion screens in mutagenized homozygous *lof* founders. Several revertant lines were recovered from ENU screens, each with independent nonsynonymous mutations within the gene encoding the potassium channel Kcnh2a (lof^{fr1} Y669N, lof^{fr2} L739Q, lof^{WL4}, and lof^{WL7} Y418X; Figures 3E and S3A). Further, a single revertant was isolated from a γ-ray screen (lof^{j6g1}), which disrupted much of the mapping interval, including removal of the sequence upstream of the *kcnh2a* transcription start site (Figure S3A).[20] Homozygous mutants of *kcnh2a* revertants have wild-type fin proportions (Figure 3E). The ability to revert the mutant phenotype by loss-of-function *kcnh2a* alleles in *cis* to *lof* indicates the mutant imparts a gain-of-function effect.

While these data suggest the function of *kcnh2a* is required for the *lof* phenotype, the causative genetic lesion underlying the mutant was unknown. Through analysis of long sequencing reads (PacBio), we identified an inversion on chromosome 2 in *lof* with break points between 5′ *prrx1a* and 5′ *kcnh2a* without affecting the coding region of either gene (Figures 3F, S3B, and S3C). This altered position of Kcnh2a in *lof* suggests that the effect may be due to enhanced or new expression of *kcnh2a* during

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

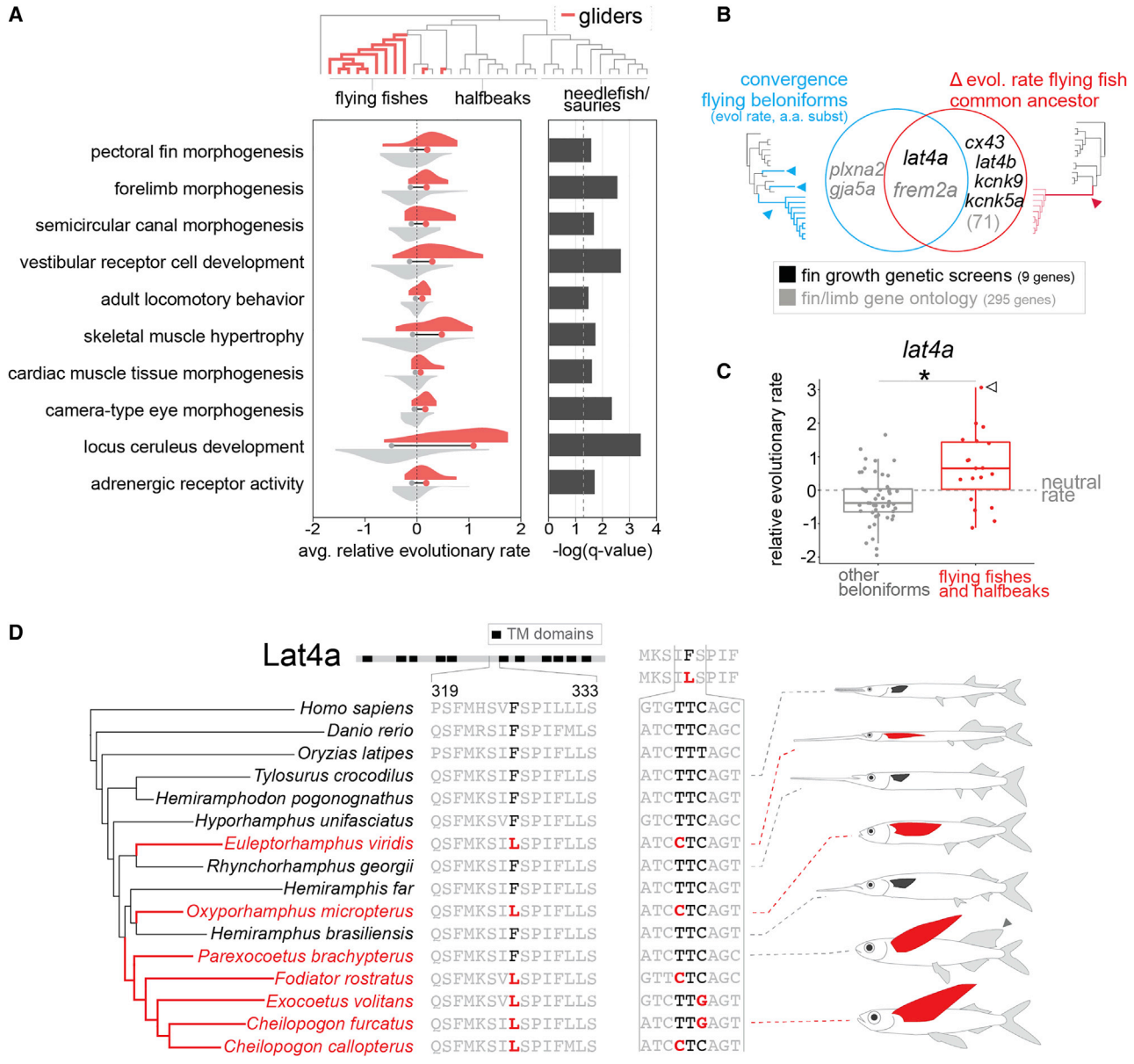

Figure 2. Convergent and ancestral genetic signatures in fin and limb genes in the evolution of gliding beloniforms

(A) Comparison of average relative evolutionary rate between the gliding and non-gliding beloniform nodes across specific gene ontology terms across the beloniform phylogeny (gliding fish lineages highlighted in red). Histograms represent the distribution across each node in the phylogeny of the average relative evolutionary rate for all genes within a given gene ontology term. The mean of the average relative evolutionary rates in gliding and non-gliding beloniform nodes is denoted at the base of each histogram. Significance (−log(q-value)) based on false discovery rate (FDR)-corrected p values from a Wilcoxon signed-rank test of differences between the mean relative evolutionary rate of gliding beloniforms compared to non-gliding beloniform branches. For full enrichment data, see Data S1F.

(B) Intersection of molecular convergence and changes to evolutionary rate in key fin-associated genes, as defined from genetic screens and gene ontology sets, at key nodes in the Beloniformes phylogeny associated with flight. See Data S1H for full gene list. We detected parallel amino acid substitutions (Data S1G) and convergent changes to relative evolutionary rate (Data S1I) in gliding beloniforms. We further identified sequence elements under accelerated or constrained evolution in the ancestral flying fish lineage (Exocoetidae; Data S1J).

(C) Elevated relative evolutionary rate in *lat4a* in gliding beloniforms. Each dot represents a node in the beloniform phylogeny. *$p < 0.05$ for two-tailed t test. White arrowhead indicates *Parexocoetus brachypterus*. Lat4a also has a phyloP acceleration $p = 0.008$ on the common ancestral node of flying fishes (Data S1J).

(D) Convergent amino acid substitution in a highly conserved region of Lat4a through multiple codon mutations in flying fishes and flying halfbeaks. This substitution is found in both lineages of flying halfbeak and all flying fishes (Exocoetidae) with the exception of the sailfin flying fish, *P. brachypterus*, which has an elongated dorsal fin (gray arrowhead).

See also Figures S1 and S2.

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

CellPress

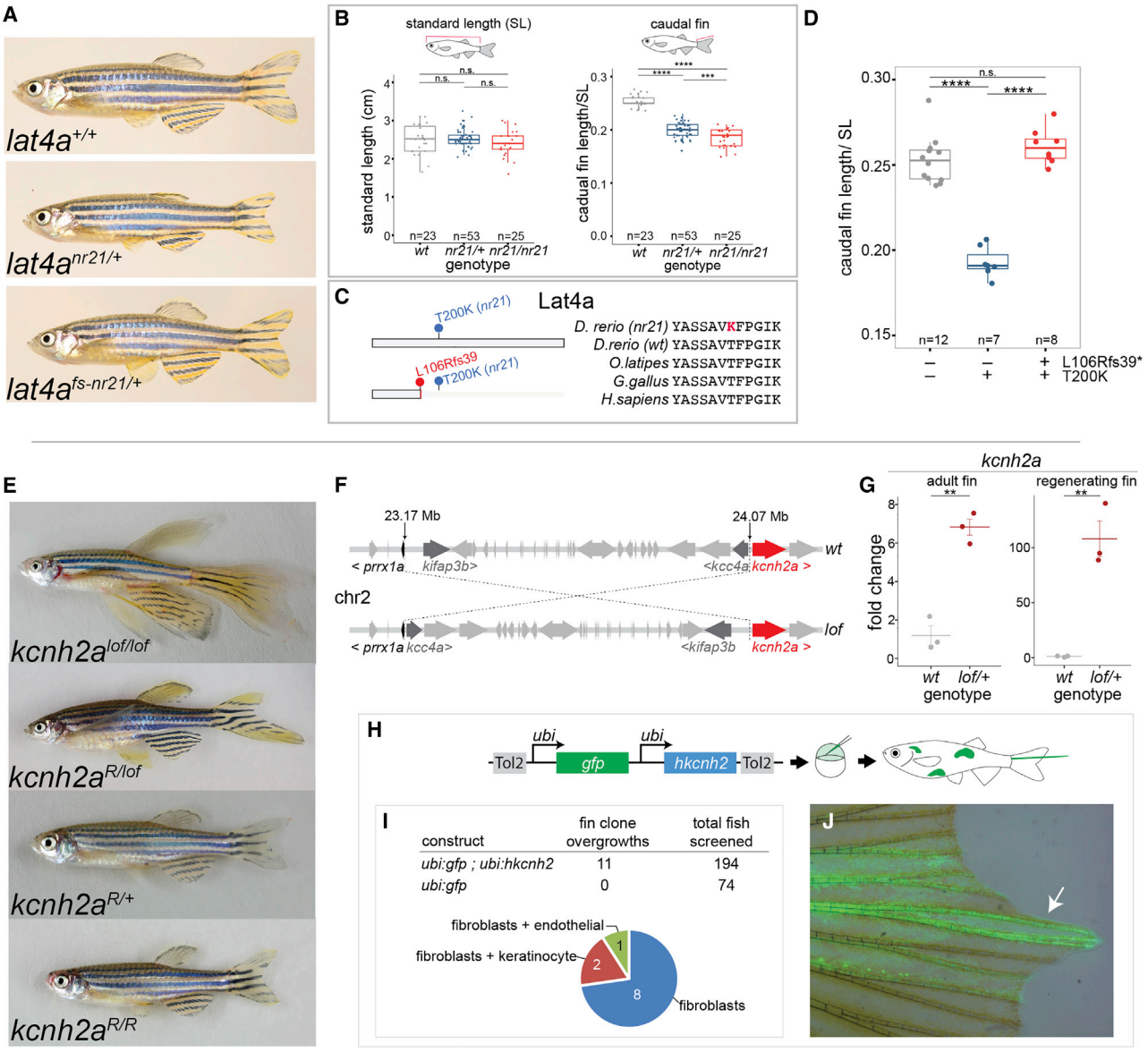

Figure 3. Identification of zebrafish mutants in potassium channels and amino acid transporters in the regulation of fin size

(A) Images of wild type, *nr21* short fin mutant, and *nr21* in *cis* frameshift revertant (L106Rfs39, *nr21*) zebrafish.

(B) Comparison of fish standard length (SL) and caudal fin length in wild type, heterozygous (*nr21*/+), and homozygous (*nr21*/*nr21*) individuals. p values generated through Tukey's HSD. ****adjusted $p \leq 0.0001$ and ***adjusted $p \leq 0.001$. n.s. indicates not significant.

(C) Multiple sequence alignment of Lat4a showing the predicted T200K substitution of *nr21* and location of frameshift reversion allele.

(D) Caudal fin length normalized to fish SL in wild type, *nr21*, and revertant mutants. Schematic indicates location of mutations on Lat4a. ****Tukey's HSD adjusted $p \leq 0.0001$.

(E) Revertant mutants of the zebrafish *longfin*^{dt2} (*lof*) mutant obtained from a mutagenesis screen. Reversion alleles mapped to loss-of-function mutations in the voltage-gated potassium channel *kcnh2a* are shown.

(F) Identified chromosomal inversion in *lof*. Note this inversion juxtaposes the regulatory region of *prrx1a* upstream of the *kcnh2a* transcription start site. For detailed positional mapping information, see Figure S3.

(G) qRT-PCR showing upregulation of *kcnh2a* in adult and regenerating caudal fins of *lof* compared to wild type. Midline represents mean; errors bars represent ±1 SEM.

(H) Mosaic overexpression assay to assess effect of *hKCNH2* overexpression on fin growth. The plasmid is injected into single-cell zebrafish embryos with Tol2 transposase mRNA and is incorporated randomly into the genome, resulting in a mosaic patchwork of gene overexpression in the adult.

(I) Number of overgrowths observed in injected zebrafish and marked cells underlying the overgrowth.

(J) Example fin overgrowth showing GFP+ fibroblast clones within the overgrown fin rays (arrow).

See also Figures S2 and S3.

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

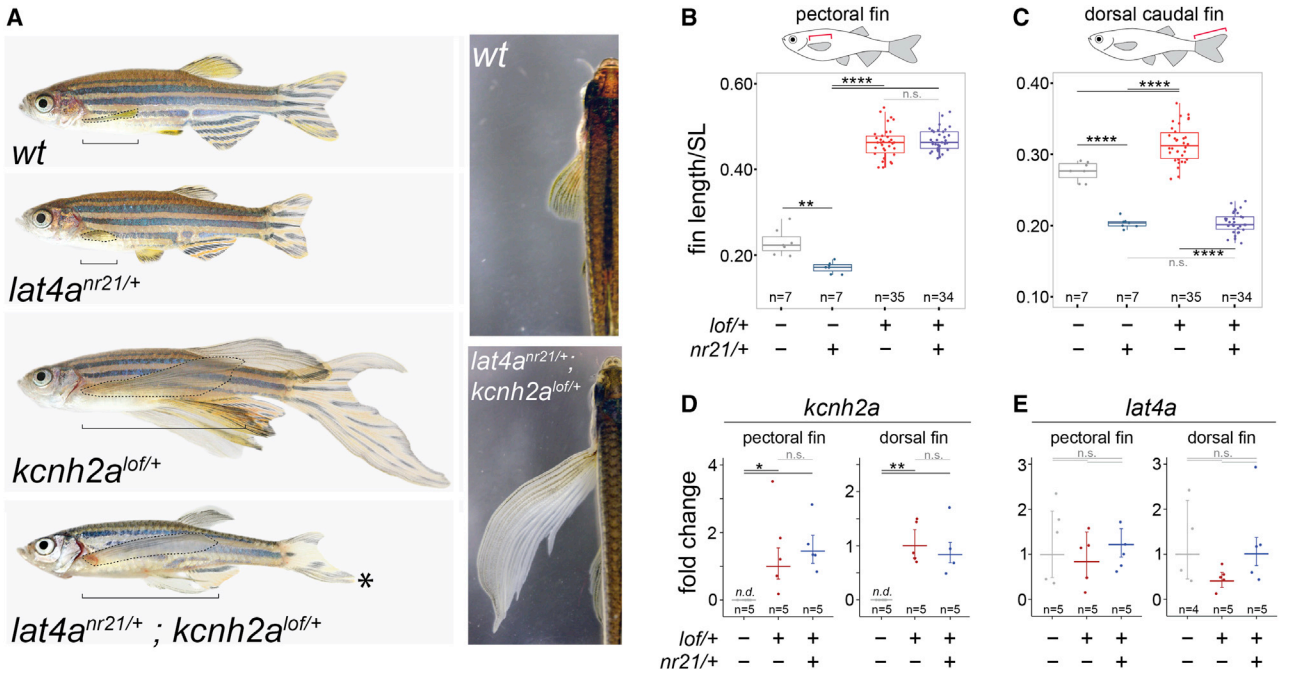

Figure 4. Intersection of bioelectric and Lat4a-mediated growth pathways is sufficient to phenocopy the evolved flying fish fin allometry

(A) Images from representative fish from a *lat4a*$^{nr21/+}$ × *kcnh2a*$^{lof/+}$ cross indicate *lat4a* is a fin suppressor of overgrowth caused by increased potassium channel function. Brackets indicate length of pectoral fin. Pectoral fins are outlined with dashed line. As in flying fishes, the *nr21/lof* trans-heterozygotes exhibit elongated paired fins and a hypocercal caudal fin. Asterisk (*) highlights hypocercal caudal fin.

(B and C) Pectoral fin length (B) and caudal fin length (C) suppression of overgrowth in medial fin by *lat4a* mutation. In (B) and (C), length normalized by SL.

(D) qRT-PCR of *kcnh2a* in pectoral and dorsal fins containing the *lof* inversion. Expression of *kcnh2a* was not detected (n.d.) in wild-type fins. Note, no significant change in expression of *kcnh2a* is observed in fins containing the *nr21* mutation. Fold change relative to *lof*/+ is shown.

(E) qRT-PCR of *lat4a* showing no significant difference in *lat4a* expression across fins and genotypes. Fold change relative to wild type is shown.

In (D) and (E), midline represents mean; errors bars represent ±1 SEM. In (B)–(E), p values are generated through Tukey's HSD. ****adjusted $p \leq 0.0001$, **adjusted $p \leq 0.01$. and **adjusted $p \leq 0.05$. n.s. indicates not significant.

development driven by elements proximal to the *prrx1a* locus. Consistent with this hypothesis, expression analysis from resting and regenerating fins reveals that the *lof* mutation results in an upregulation of *kcnh2a* in fins (Figure 3G). Upregulation of human *KCNH2* in fin mesenchyme during development (*ubi:hKCNH2;ubi:GFP*) was sufficient to cause localized overgrowth of rays autonomous to the fin clone (Figures 3H–3J). Thus, the chromosomal rearrangement in *lof* changes the regulation of the potassium channel *kcnh2a*, enhancing *kcnh2a* expression in the forming fin and is sufficient for increased fin size. These findings are supported by a recent paper detailing comparable analyses of *kcnh2a*^{lof} in mediating size of the fin, though without identification of the genetic lesion underlying the *lof* phenotype.[21]

Convergent genomic signatures of fin allometry

With the broadened complement of genes known to affect fin proportion,[22] we compared sets of genes affecting fin proportion with patterns of genomic variation in gliding beloniforms. We first identified convergent signals shared among species having aerial gliding behavior. Compared with 14–28 convergent amino acid substitutions in control comparisons (Figure S1C), we identified 44 amino acid substitutions that are present only within all three lineages of gliding beloniforms (Figures 2B and S1C; Data S1G). We then intersected the identified convergent amino acid substitutions with genes defined from forward genetic screens in zebrafish and from studies of fin and limb development (Figure 2B; Data S1H). Intriguingly, we detected a shared amino acid substitution in *lat4a* that is found across multiple lineages having enlarged pectoral fins (F326L; Figures 2B–2D; Data S1G). This amino acid substitution is caused by multiple DNA mutations (CTC and TTG), suggesting independent derivation in evolution (Figure 2D). The multiple F326L codon alleles (CTC and TTG) do not cluster in either a *lat4a* phylogenetic gene tree or in a localized phylogenetic tree from *lat4a* exon 8 (Figures S1D and S1E). Further, we did not detect informative SNPs flanking the F326L alleles that would indicate the presence of a larger ancestral haplotype. While these data support a model of convergent origins of the F326L mutation in gliding beloniforms, we cannot exclude incomplete lineage sorting or introgression of ancestral F326L haplotypes specifically within gliding lineages. The *lat4a* locus not only harbors shared amino acid substitutions among gliding fishes of Beloniformes but has an accelerated rate of evolution as well (Figure 2C). Intriguingly, only one flying fish species in our dataset lacks the convergent *lat4a* allele, the sailfin flying fish *Parexocoetus brachypterus*, unique among the sequenced flying fishes in having an elongated dorsal fin (Figure 2D). The evolutionary rate of *P. brachypterus lat4a* is highly accelerated and is distinct from that of other flying fishes (Figure 2C), revealing unique evolution of this gene in this species.

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

In addition to analysis of convergent amino acid substitutions, we explored convergent shifts in evolutionary rate in flying fishes. We identify 48 genes (out of 17,937 analyzed) as having convergent evolutionary rate shifts after false discovery rate correction, though few of these genes are known to be involved in appendage development (Figure 2B; Data S1H and S1I). Of note, the potassium channel *kcnj13* shows convergent acceleration in gliding beloniforms though is not significant after multiple hypothesis correction. Overexpression of *kcnj13* in zebrafish, like *kcnk5b* and *kcnh2a* (Figures 3H–3J), is sufficient to induce fin overgrowth.[23] Though many of the genes in the set are not directly attributed to size regulation based on current knowledge, these genes may serve as a foundation for future discovery.

Genomic footprints of flying fish evolution

In addition to analyses of molecular convergence in gliding beloniforms, we explored patterns of accelerated or constrained evolution within the common ancestor of all flying fishes, where the shared elongated pectoral fin allometry is likely to have first emerged. Several genes involved in fin and limb development are under varied selection regimes, including *lat4a* (Figure 2B; Data S1J). There is also accelerated sequence evolution of *kcnk9*, overexpression of which is sufficient to drive fin overgrowth in zebrafish,[23] and in *kcnk5a*, the paralog of *kcnk5b* altered in the long-finned zebrafish mutant, *another longfin* (Figures 2B and S1A; Data S1J).[19] These data are in agreement with accumulating evidence both in experimental mutant analyses as well as association in non-model species that potassium channels are key regulators of fin proportion in teleost fishes.[24,25]

Further genomic analysis identified accelerated sequence evolution in the flying fish common ancestor within the transcription factor *evx1*, which regulates segmentation of fin lepidotrichia (Data S1J).[26] Intriguingly, *evx1* also shows subtle convergence in flying halfbeaks (Data S1I). Additionally, the gap junction protein *cx43*, mutations in which cause short fins in zebrafish,[27,28] is under constrained sequence evolution along the ancestral branch of flying fishes (Figure 2B; Data S1J). Retinoic acid is involved in outgrowth and patterning of limbs and is active along a proximal-distal gradient.[29] Interestingly, the retinoic-acid-inactivating enzymes *cyp26b1* and *cyp27c1* are both under accelerated sequence evolution in our dataset (Data S1J). Consistent with these data, we observed constraint in the evolution of the retinoic acid synthesis gene *aldh1a2* in gliding beloniforms (Data S1I), though this signature is not significant after false discovery rate correction.

Regulatory changes near fin and limb developmental genes

We assigned CNEs to neighboring genes to assess whether there is selection in putative gene-regulatory regions in the common ancestor of flying fishes (GREAT algorithm).[30] We find a significant enrichment for accelerated sequence evolution in CNEs near multiple gene classes associated with gliding behavior and morphology (Figures S4A–S4D; Data S1K–S1N), specifically genes associated with pectoral fin and appendage development (Figure S4B; Data S1K). A particularly intriguing locus is *sall1a*. *Sall1a* is required for pectoral fin development in zebrafish,[31] limb development in mouse,[32] and is implicated in the reduction of emu forelimbs.[33] There is widespread accelerated sequence evolution at the *sall1a* locus along the ancestral flying fish branch, both in coding and non-coding regions (Figure S4E; Data S1J and S1K). Intriguingly, the *sall1a* locus also exhibits a convergent shift in evolutionary rate in the gliding beloniforms though is not significant after false discovery rate correction (Data S1I).

Notably, our CNE analysis also was congruent with mutational analyses in the zebrafish. Specifically, in the common ancestor of flying fishes, we see enrichment for accelerated sequence evolution in CNEs near genes involved in the "cellular response to potassium ion" and near genes involved in "cellular response to amino acids," specifically leucine, one of the amino acids transported by Lat4 (Figure S4B; Data S1K).[16]

Sufficiency of simple genetic changes to phenocopy flying fish morphology in zebrafish

Given comparative signatures in Beloniformes in genes and in CNEs that are involved in potassium and amino acid transport, we asked how these genetic mechanisms may interact to regulate fin size and patterning. We made crosses between zebrafish mutants affecting potassium channel regulation and *lat4a*. Interestingly, double mutants between *kcnh2a*$^{lof/+}$ and *lat4a*$^{nr21/+}$ are viable and show specific, non-additive variation in proportion among fins: while all fins of *kcnh2a*$^{lof/+}$ are overgrown and all fins of *lat4a*$^{nr21/+}$ are shortened, the *lat4a*$^{nr21/+}$;*kcnh2a*$^{lof/+}$ fish show wild-type-sized medial fins, while the paired fins are overgrown (Figure 4). Thus, *lat4a*$^{nr21/+}$ is a medial fin suppressor of *kcnh2a*$^{lof/+}$, exposing fin-type-specific regulation of size. Surprisingly, the caudal fin of *kcnh2a*$^{lof/+}$;*lat4a*$^{nr21/+}$, while generally wild type in size, is hypocercal, with an elongated ventral lobe. The overall fin phenotype of the *kcnh2a*^{lof};*lat4a*^{nr21} fish is thus remarkably similar to that of flying fishes, suggesting the interaction between these two genes is sufficient to phenocopy this key innovation and organismal bauplan of flying fishes (Figure 4).

As ectopic expression of *kcnh2a* is sufficient to drive fin overgrowth (Figure 3), we hypothesized that the fin-specific suppression of *lof* by *nr21* might be linked to modulation of *kcnh2a* expression levels. Surprisingly, qRT-PCR revealed that *kcnh2a* expression does not vary significantly in pectoral or dorsal fins containing the *nr21* mutation (Figures 4D and 4E). Thus, the mechanism of fin-specific suppression of *lof* by *lat4a* mutations is occurring downstream of bioelectric signaling signals.

Developmental constraint and potential

The combination of forward genetic approaches, both experimental and evolutionary, is a powerful means to parse out regulatory pathways of development and physiology. The detection of comparative genomic signatures associated with bioelectric signaling and amino acid transport in flying fishes and functional analyses in distantly related zebrafish points to a role for these genes in the development and evolution of allometry. There appear to be varied means of impacting bioelectric signaling and its effect on fin proportion, ranging from point mutations to regulation of gene expression. That simple genetic changes modulating this signaling can lead to coordinated alteration in fin patterning may underlie the general evolvability and diversity of fin allometry. Aerial gliding in fishes has evolved several times independently in teleosts,[1] and multiple examples of the "flying fish" morphotype can be seen in the fossil record.[34,35] That such a bauplan has evolved multiple times may represent a developmental bias in shaping morphology available for selection.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Zebrafish husbandry and lines
- METHOD DETAILS
  - Chemical and gamma-ray mutagenesis
  - Mapping and mutational analysis of lof (dt2)
  - Overexpression constructs
  - Isolation and mapping of nr21 mutant line
  - CRISPR-Cas9 gRNA design and injection
  - qRT-PCR
  - Targeted Sequence Capture Design
  - Specimen tissue collection and sequencing library preparation
  - Reference contig assembly
  - Identification of orthologs
  - Read coverage and depth of targeted regions
  - Nucleotide diversity and recovery of population variation
  - Multiple sequence alignment
  - Beloniformes phylogeny
  - Reconstruction of gene sequences
  - CNE association with genes
  - Ontology annotations
  - Evolutionary sequence rate analysis
  - Analysis of molecular convergence
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.08.054.

ACKNOWLEDGMENTS

This work was supported by a John Simon Guggenheim Fellowship, Milton Foundation Fund, and partially supported by NIH grant R01HD084985 to M.P.H.; a National Science Foundation Doctoral Dissertation Improvement Grant (DDIG) to J.M.D. (DEB-1407092); a German Research Foundation fellowship to N.B. (BL 1614/1-1); and an NSERC Discovery Grant to N.R.L. Specimen and tissue collection and curation was made possible by generous assistance from R. Pitman, L. Ballance, B. Collette, H.H. Tan, E. Lewallen, D. Bloom, M. Sabaj, J.A. Dorman, NOAA, the Scripps Institution of Oceanography, and the Royal Ontario Museum. The authors would like to acknowledge assistance by the Central Caribbean Marine Institute (CCMI) and their generous support of this research program. This work is dedicated to the insight and career of Dr. Steven Johnson, who helped shape and identify the novel role of *kcnh2a* in regulating the *longfin* phenotype.

AUTHOR CONTRIBUTIONS

J.M.D. and M.P.H. designed the project. Cloning of the zebrafish mutant *lof* was performed by M.K.I., C.W.H., S.L.J., J.M.D., and M.P.H. Cloning of the *nr21* zebrafish mutant was performed by J.M.D. qRT-PCR was performed by H.B., J.L., and J.M.D. *lof* overexpression experiments and analysis of *nr21* fin size were performed by N.B. Beloniform tissue samples were provided by N.R.L. Genomic analysis was performed by J.M.D. The manuscript was written by J.M.D. and M.P.H. with comments from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 5, 2021
Revised: July 9, 2021
Accepted: August 20, 2021
Published: September 16, 2021

REFERENCES

1. Davenport, J. (1994). How and why do flying fish fly? Rev. Fish Biol. Fish. *4*, 184–214.
2. Lewallen, E.A., Pitman, R.L., Kjartanson, S.L., and Lovejoy, N.R. (2011). Molecular systematics of flyingfishes (Teleostei: Exocoetidae): evolution in the epipelagic zone. Biol. J. Linn. Soc. Lond. *102*, 161–174.
3. Dasilao, J.C., Sasaki, K., and Okamura, O. (1997). The hemiramphid, *Oxyporhamphus*, is a flyingfish (exocoetidae). Ichthyol. Res. *44*, 101–107.
4. Lovejoy, N.R., Iranpour, M., and Collette, B.B. (2004). Phylogeny and jaw ontogeny of beloniform fishes. Integr. Comp. Biol. *44*, 366–377.
5. Hughes, L.C., Ortí, G., Huang, Y., Sun, Y., Baldwin, C.C., Thompson, A.W., Arcila, D., Betancur-R, R., Li, C., Becker, L., et al. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. USA *115*, 6249–6254.
6. Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T.J., Coll, M., and Alfaro, M.E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. Nature *559*, 392–395.
7. Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., and Ortí, G. (2017). Phylogenetic classification of bony fishes. BMC Evol. Biol. *17*, 162.
8. Daane, J.M., Rohner, N., Konstantinidis, P., Djuranovic, S., and Harris, M.P. (2016). Parallelism and epistasis in skeletal evolution identified through use of phylogenomic mapping strategies. Mol. Biol. Evol. *33*, 162–173.
9. Daane, J.M., Dornburg, A., Smits, P., MacGuigan, D.J., Brent Hawkins, M., Near, T.J., William Detrich Iii, H., and Harris, M.P. (2019). Historical contingency shapes adaptive radiation in Antarctic fishes. Nat. Ecol. Evol. *3*, 1102–1109.
10. Lovejoy, N.R. (2000). Reinterpreting recapitulation: systematics of needlefishes and their allies (Teleostei: Beloniformes). Evolution *54*, 1349–1362.
11. Kasumyan, A.O. (2004). The vestibular system and sense of equilibrium in fish. J. Ichthyol. *44*, 224–268.
12. Dasilao, J.C., and Sasaki, K. (1998). Phylogeny of the flyingfish family Exocoetidae (Teleostei, Beloniformes). Ichthyol. Res. *45*, 347–353.
13. Baylor, E.R. (1967). Air and water vision of the Atlantic flying fish, Cypselurus heterurus. Nature *214*, 307–309.
14. Morris, L.S., McCall, J.G., Charney, D.S., and Murrough, J.W. (2020). The role of the locus coeruleus in the generation of pathological anxiety. Brain Neurosci. Adv. *4*, 2398212820930321.
15. Bowen, M.E., Henke, K., Siegfried, K.R., Warman, M.L., and Harris, M.P. (2012). Efficient mapping and cloning of mutations in zebrafish by low-coverage whole-genome sequencing. Genetics *190*, 1017–1024.
16. Bodoy, S., Martín, L., Zorzano, A., Palacín, M., Estévez, R., and Bertran, J. (2005). Identification of LAT4, a novel amino acid transporter with system L activity. J. Biol. Chem. *280*, 12002–12011.
17. Guetg, A., Mariotta, L., Bock, L., Herzog, B., Fingerhut, R., Camargo, S.M.R., and Verrey, F. (2015). Essential amino acid transporter Lat4 (Slc43a2) is required for mouse development. J. Physiol. *593*, 1273–1289.

18. van Eeden, F.J., Granato, M., Schach, U., Brand, M., Furutani-Seiki, M., Haffter, P., Hammerschmidt, M., Heisenberg, C.P., Jiang, Y.J., Kane, D.A., et al. (1996). Genetic analysis of fin formation in the zebrafish, Danio rerio. Development *123*, 255–262.

19. Perathoner, S., Daane, J.M., Henrion, U., Seebohm, G., Higdon, C.W., Johnson, S.L., Nüsslein-Volhard, C., and Harris, M.P. (2014). Bioelectric signaling regulates size in zebrafish fins. PLoS Genet. *10*, e1004080.

20. Iovine, M.K., and Johnson, S.L. (2002). A genetic, deletion, physical, and human homology map of the long fin region on zebrafish linkage group 2. Genomics *79*, 756–759.

21. Stewart, S., Le Bleu, H.K., Yette, G.A., Henner, A.L., Robbins, A.E., Braunstein, J.A., and Stankunas, K. (2021). longfin causes cis-ectopic expression of the kcnh2a ether-a-go-go K+ channel to autonomously pro-long fin outgrowth. Development *148*, dev199384.

22. Harris, M.P., Daane, J.M., and Lanni, J. (2021). Through veiled mirrors: fish fins giving insight into size regulation. WIREs Dev. Biol. *10*, e381.

23. Silic, M.R., Wu, Q., Kim, B.H., Golling, G., Chen, K.H., Freitas, R., Chubykin, A.A., Mittal, S.K., and Zhang, G. (2020). Potassium channel-associated bioelectricity of the dermomyotome determines fin patterning in Zebrafish. Genetics *215*, 1067–1084.

24. Schartl, M., Kneitz, S., Ormanns, J., Schmidt, C., Anderson, J.L., Amores, A., Catchen, J., Wilson, C., Geiger, D., Du, K., et al. (2021). The develop-mental and genetic architecture of the sexually selected male ornament of swordtails. Curr. Biol. *31*, 911–922.e4.

25. Kon, T., Omori, Y., Fukuta, K., Wada, H., Watanabe, M., Chen, Z., Iwasaki, M., Mishina, T., Matsuzaki, S.S., Yoshihara, D., et al. (2020). The genetic basis of morphological diversity in domesticated goldfish. Curr. Biol. *30*, 2260–2274.e6.

26. Schulte, C.J., Allen, C., England, S.J., Juárez-Morales, J.L., and Lewis, K.E. (2011). Evx1 is required for joint formation in zebrafish fin dermoske-leton. Dev. Dyn. *240*, 1240–1248.

27. Iovine, M.K., Higgins, E.P., Hindes, A., Coblitz, B., and Johnson, S.L. (2005). Mutations in connexin43 (GJA1) perturb bone growth in zebrafish fins. Dev. Biol. *278*, 208–219.

28. Sims, K., Jr., Eble, D.M., and Iovine, M.K. (2009). Connexin43 regulates joint location in zebrafish fins. Dev. Biol. *327*, 410–418.

29. Maden, M. (1982). Vitamin A and pattern formation in the regenerating limb. Nature *295*, 672–675.

30. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional inter-pretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

31. Harvey, S.A., and Logan, M.P.O. (2006). sall4 acts downstream of tbx5 and is required for pectoral fin outgrowth. Development *133*, 1165–1173.

32. Kawakami, Y., Uchiyama, Y., Rodriguez Esteban, C., Inenaga, T., Koyano-Nakagawa, N., Kawakami, H., Marti, M., Kmita, M., Monaghan-Nichols, P., Nishinakamura, R., and Izpisua Belmonte, J.C. (2009). Sall genes regulate region-specific morphogenesis in the mouse limb by modulating Hox ac-tivities. Development *136*, 585–594.

33. Young, J.J., Grayson, P., Edwards, S.V., and Tabin, C.J. (2019). Attenuated Fgf signaling underlies the forelimb heterochrony in the Emu Dromaius novaehollandiae. Curr. Biol. *29*, 3681–3691.e5.

34. Xu, G.H., Zhao, L.J., Gao, K.Q., and Wu, F.X. (2013). A new stem-neo-pterygian fish from the Middle Triassic of China shows the earliest over-water gliding strategy of the vertebrates. Proc. Biol. Sci. *280*, 20122261.

35. Tintori, A., and Sassi, D. (1992). *Thoracopterus* Bronn (Osteichthyes: Actinopterygii): a gliding fish from the Upper Triassic of Europe. J. Vertebr. Paleontol. *12*, 265–283.

36. Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature *447*, 714–719.

37. Warren, W.C., García-Pérez, R., Xu, S., Lampert, K.P., Chalopin, D., Stöck, M., Loewe, L., Lu, Y., Kuderna, L., Minx, P., et al. (2018). Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. Nat. Ecol. Evol. *2*, 669–679.

38. Schartl, M., Walter, R.B., Shen, Y., Garcia, T., Catchen, J., Amores, A., Braasch, I., Chalopin, D., Volff, J.-N., Lesch, K.-P., et al. (2013). The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nat. Genet. *45*, 567–572.

39. Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature *513*, 375–381.

40. Haffter, P., Odenthal, J., Mullins, M.C., Lin, S., Farrell, M.J., Vogelsang, E., Haas, F., Brand, M., van Eeden, F.J.M., Furutani-Seiki, M., et al. (1996). Mutations affecting pigmentation and shape of the adult zebrafish. Dev. Genes Evol. *206*, 260–276.

41. Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS ONE *9*, e98186.

42. Balciunas, D., Wangensteen, K.J., Wilber, A., Bell, J., Geurts, A., Sivasubbu, S., Wang, X., Hackett, P.B., Largaespada, D.A., McIvor, R.S., and Ekker, S.C. (2006). Harnessing a high cargo-capacity trans-poson for genetic applications in vertebrates. PLoS Genet. *2*, e169.

43. Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., and Chien, C.-B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev. Dyn. *236*, 3088–3099.

44. Mosimann, C., Kaufman, C.K., Li, P., Pugach, E.K., Tamplin, O.J., and Zon, L.I. (2011). Ubiquitous transgene expression and Cre-based recom-bination driven by the ubiquitin promoter in zebrafish. Development *138*, 169–177.

45. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.

46. Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. Brief. Bioinform. *12*, 41–51.

47. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

48. Kowalczyk, A., Meyer, W.K., Partha, R., Mao, W., Clark, N.L., and Chikina, M. (2019). RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics *35*, 4815–4817.

49. Sedlazeck, F.J., Rescheneder, P., and von Haeseler, A. (2013). NextGenMap: fast and accurate read mapping in highly polymorphic ge-nomes. Bioinformatics *29*, 2790–2791.

50. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

52. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

53. Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., and Schlötterer, C. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS ONE *6*, e15925.

54. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

55. Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E.J.P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS ONE *6*, e22594.

56. Borowiec, M.L. (2019). Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. J. Open Source Softw. *4*, 1635.

57. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19* (*Suppl 6*), 153.

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

59. Huang, X., and Madan, A. (1999). CAP3: a DNA sequence assembly program. Genome Res. *9*, 868–877.

60. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

61. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

62. Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) *2011*, bar030.

63. Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, W401–W407.

64. Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. *44* (W1), W272-6.

65. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

66. Chen, K., Durand, D., and Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. J. Comput. Biol. *7*, 429–447.

67. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS ONE *7*, e46688.

68. Nüsslein-Volhard, C., and Dahm, R. (2002). Zebrafish: A Practical Approach (Oxford University).

69. Rohner, N., Perathoner, S., Frohnhöfer, H.G., and Harris, M.P. (2011). Enhancing the efficiency of N-ethyl-N-nitrosourea-induced mutagenesis in the zebrafish. Zebrafish *8*, 119–123.

70. Fritz, A., Rozowski, M., Walker, C., and Westerfield, M. (1996). Identification of selected gamma-ray induced deficiencies in zebrafish using multiplex polymerase chain reaction. Genetics *144*, 1735–1745.

71. Bowen, M.E., Boyden, E.D., Holm, I.A., Campos-Xavier, B., Bonafé, L., Superti-Furga, A., Ikegawa, S., Cormier-Daire, V., Bovée, J.V., Pansuriya, T.C., et al. (2011). Loss-of-function mutations in PTPN11 cause metachondromatosis, but not Ollier disease or Maffucci syndrome. PLoS Genet. *7*, e1002050.

72. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

73. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.-R.J. (2013). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS ONE *8*, e68708.

74. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

75. Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M.J., Amode, R., Brent, S., et al. (2016). Ensembl comparative genomics resources. Database (Oxford) *2016*, bav096.

76. Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. *39*, D152–D157.

77. Dimitrieva, S., and Bucher, P. (2013). UCNEbase–a database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res. *41*, D101–D109.

78. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

79. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

80. Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. *33*, 1654–1668.

81. Daane, J.M., Auvinet, J., Stoebenau, A., Yergeau, D., Harris, M.P., and Detrich, H.W., 3rd. (2020). Developmental constraint shaped genome evolution and erythrocyte loss in Antarctic fishes following paleoclimate change. PLoS Genet. *16*, e1009173.

82. Daub, J.T., Moretti, S., Davydov, I.I., Excoffier, L., and Robinson-Rechavi, M. (2017). Detection of pathways affected by positive selection in primate lineages ancestral to humans. Mol. Biol. Evol. *34*, 1391–1402.

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
TOP10 Competent Cells	Thermo Fischer Scientific	Cat# C404010
Chemicals, peptides, and recombinant proteins		
N-ethyl-N-nitrosourea (ENU)	Sigma	Cat# N3385
SeqCap EZ Developer Reagent	Roche	Cat# 06684335001
KAPA HyperPrep Kit	Roche	Cat# 07137923001
Agencourt AMPure XP	Beckman Coulter	Cat# A63881
QIAGEN DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
TRIzol Reagent	Thermo Fischer Scientific	Cat# 15596026
SuperScript III Reverse Transcription Kit	Thermo Fischer Scientific	Cat# 18080093
Power SYBR Green Master Mix	Applied Biosystems	Cat# 4367659
Cas9 SmartNuclease mRNA	SBI System Biosciences	Cat# CAS500A-1
Ambion MEGAscript SP6 Transcription kit	Thermo Fischer Scientific	Cat# AM1330
SureSelect 1M Capture Array	Agilent	Cat# 06471684001
SeqCap EZ Developer Probes	Roche	Cat# 06471684001
T7 Endonuclease	New England Biolabs	Cat# M0302S
NEB Buffer 2	New England Biolabs	Cat# B7002S
ProtoScript II First Strand cDNA Synthesis Kit	New England Biolabs	Cat# E6560S
TURBO DNA-Free Kit	Thermo Fischer Scientific	Cat# AM1907
Deposited data		
longfin PacBio sequencing reads	N/A	SRA: ERX1428166
Beloniformes targeted sequencing reads	This paper	BioProject: PRJNA743939
Beloniformes assembled contigs	This paper	Zenodo: 10.5281/zenodo.5082978
Zebrafish reference genome (Zv11)	N/A	http://useast.ensembl.org/Danio_rerio/Info/Index
Medaka reference genome (MEDAKA1)	Kasahara et al.[36]	http://jul2018.archive.ensembl.org/Oryzias_latipes/Info/Index
Platyfish reference genome (Poecilia_formosa-5.1.2)	Warren et al.[37]	http://useast.ensembl.org/Xiphophorus_maculatus/Info/Index
Molly reference genome (Xipmac4.4.2)	Schartl et al.[38]	http://useast.ensembl.org/Poecilia_formosa/Info/Index
Tilapia reference genome (Orenil1.0)	Brawand et al.[39]	http://sep2019.archive.ensembl.org/Oreochromis_niloticus/Info/Index
Experimental models: organisms/strains		
lof^{j6e1} (lof$^{Df(Chr02:csnk1g2a,rnf2,kifap3b)j6g1/j6g1}$)	Iovine and Johnson[20]	ZIRC repository: ZL1494
lat4^{nr21}	This paper	*dmh26*
lat4a^{del}	This paper	*mh152*
lof^{dt2}	van Eeden et al.[18] and Haffter et al.[40]	ZIRC repository: ZL86
Oligonucleotides		
gRNA universal constant oligo: 5′- AAAA GCACCGACTCGGTGCCACTTTTTCAA GTTGATAACGGACTAGCCTTATTTTAAC TTGCTATTTCTAGCTCTAAAAC-3′	Gagnon et al.[41]	N/A
Lat4a gRNA #1 5′-‘ATTTAGGTGACA CTATAGGCCCTGTACCGTTACCTGG GTTTTAGAGCTAGAAATAGCAAG-3′	This paper	N/A

(*Continued on next page*)

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Lat4a gRNA #2 5′- ATTTAGGTGACA CTATAGGTGAATGCCACAAGACTTG GTTTTAGAGCTAGAAATAGCAAG-3′	This paper	N/A
Lat4a gRNA #3 5′- ATTTAGGTGACACTAT AGGATGCCACAAGACTTGAGG GTTTTAGAGCTAGAAATAGCAAG-3′	This paper	N/A
Lat4a gRNA #4 5′- ATTTAGGTGAC ACTATAGGCCACAAGACTTGAGGAGG GTTTTAGAGCTAGAAATAGCAAG-3′	This paper	N/A
Lat4a gRNA #5 5′- ATTTAGGTGACACTATA GGAGTGTGATGGCGCTCAGG GTTTTAGAGCTAGAAATAGCAAG-3′	This paper	N/A
kcnh2a qPCR primers: 5′- CCTCATCT TCAGTTCCCTCTAAC-3′ and 5′- GCTG CCTGTGCATCATTTC-3′	This paper	N/A
gapdh qPCR primers: 5′- GGACAC AACCAAATCAGGCATA-3′ and 5′- CGCCTTCTGCCTTAACCTCA-3′	This paper	N/A
β-actin primers: 5′-CCCAAAGCCAACAGA GAGAA-3′ and 5′-ACCAGAAGCGTAC AGAGAGA-3′	This paper	N/A
lat4a primers: 5′-ACGCAGGACG GCAATAAA-3′, 5′-GCTGAAGATACT CCGCATGAA-3′	This paper	N/A
Recombinant DNA		
pminiTol2	Balciunas et al.[42]	Addgene plasmid #31829
pME-EGFP	Kwan et al.[43]	Tol2kit #383
pENTR5′_ubi	Mosimann et al.[44]	Addgene plasmid #27320
Software and algorithms		
PHAST v1.4	Pollard et al.[45] and Hubisz et al.[46]	http://compgen.cshl.edu/phast/
statsmodels v.0.6.1	N/A	https://www.statsmodels.org/stable/index.html
IQTree v1.7beta2	Nguyen et al.[47]	http://www.iqtree.org/
RERConverge v0.1.0	Kowalczyk et al.[48]	https://github.com/nclark-lab/RERconverge
Biopython v.1.70	N/A	https://biopython.org/
NextGenMap v0.5.5	Sedlazeck et al.[49]	https://cibiv.github.io/NextGenMap/
BEDTools v2.23.0	Quinlan and Hall[50]	https://bedtools.readthedocs.io/en/latest/
SAMtools v1.9	Li et al.[51]	http://www.htslib.org/
Integrative Genome Viewer (IGV) v2.8.9	Robinson et al.[52]	https://software.broadinstitute.org/software/igv/
SciPy v.0.18.1	N/A	https://www.scipy.org/
PoPoolation v1.2.2	Kofler et al.[53]	https://sourceforge.net/p/popoolation/wiki/Main/
MAFFT v 7.313	Katoh and Standley[54]	https://mafft.cbrc.jp/alignment/software/
MACSE v2.03	Ranwez et al.[55]	https://bioweb.supagro.inra.fr/macse/
Spruceup v 2020.2.19	Borowiec[56]	https://github.com/marekborowiec/spruceup
ASTRAL v5.6.1	Zhang et al.[57]	https://github.com/smirarab/ASTRAL
FASTX toolkit	N/A	http://hannonlab.cshl.edu/fastx_toolkit
Trimmomatic v.0.36	Bolger et al.[58]	https://github.com/usadellab/Trimmomatic
CAP3 v 02/10/15	Huang and Madan[59]	http://doua.prabi.fr/software/cap3

(*Continued on next page*)

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
usearch-6.1	Edgar[60]	https://drive5.com/usearch/manual/uclust_algo.html
ncbi-blast-2.2.30+	Altschul et al.[61]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Ensembl Biomart	Kinsella et al.[62]	http://useast.ensembl.org/biomart/martview//61e23c69ebbfd853fc12e7bb6975ae34
ChopChop	Montague et al.[63] and Labun et al.[64]	https://chopchop.cbu.uib.no/
Novoalign	N/A	http://www.novocraft.com/),
Canu v. r10117	Koren et al.[65]	https://github.com/marbl/canu
Notung-2.9	Chen et al.[66]	http://www.cs.cmu.edu/~durand/Notung/
PROVEAN	Choi et al.[67]	http://provean.jcvi.org/index.php

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jacob Daane (j.daane@northeastern.edu).

Materials availability

Zebrafish lines generated by the authors will be distributed upon request to other researchers. Sperm from these lines have been deposited with Cryogenetics (https://www.cryogenetics.com).

Data and code availability

Raw targeted sequencing reads have been deposited in the NCBI Sequence Read Archive. Assembled reference contigs and annotations have been deposited on Zenodo. Accession numbers and DOI are listed in the Key resources table and are publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Zebrafish husbandry and lines

Zebrafish were housed and maintained as described in Nüsslein-Volhard and Dahm[68] and performed in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines at Boston Children's Hospital and Washington University in Saint Louis. A description of the husbandry and environmental conditions in housing for the fish used in these experiments at Boston Children's Hospital is available as a collection in protocols.io (https://doi.org/10.17504/protocols.io.mrjc54n). Similar conditions were present at Washington University Medical School (SLJ lab). For all experiments, adult stages were defined by reproductively mature fish $\geq$ 3 months old. Males and females were used together in analyses as both sexes showed comparable overgrowth and growth regulation. *Longfin* alleles used in this work are lof^{dt2}, the revertants *lof* (*lfr1*(Y669N), *lfr2*(L739Q), and *WL7* (Y418X). These lines are no longer maintained nor frozen, thus are used for confirmation of mapping only. The deletion line lof^{i6e1} ($lof^{Df(Chr02:csnk1g2a,rnf2,kifap3b)j6g1/j6g1}$) is present at ZIRC repository (ZL1494).[20] Lat4a alleles identified and used in this work are $lat4^{nr21}$ (*dmh26*), and $lat4a^{del}$ (*mh152*).

METHOD DETAILS

Chemical and gamma-ray mutagenesis

For identification of lof^{dt2} revertants, homozygous lof^{dt2} mutant males were mutagenized with 3.3μM N-ethyl-N-nitrosourea (ENU) for four repeated doses following Rohner et al.[69] Founder mutagenized males were crossed to Tübingen wild-type females. F1 progeny were scored for a reduction or increase in the fin length compared to heterozygous *lof* fish. Radiation induced mutations were generated by exposing sperm from lof^{dt2} homozygous males to 250 rads of gamma-irradiation.[20,70] Irradiated sperm was used to fertilize eggs from C32 females. Progeny exhibiting wild-type phenotypes were recovered as potential lof^{dt2} revertants. Identification and description of lof^{i6e1} is described in Iovine and Johnson.[20]

Mapping and mutational analysis of lof (dt2)

The lof^{dt2} mutant was independently mapped in both the MPH and SLJ laboratories using bulk segregant analysis and analysis of both simple sequence length polymorphisms (SSLP) and single nucleotide polymorphisms (SNPs).[20] As both datasets were consistent, the

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

mapping data were combined to refine the identified mapping interval on chromosome 2 (Figure S3A). As a complementary approach, both MPH and SLJ laboratories performed revertant screens, as described above, and identified independent nonsynonymous mutations in *kcnh2a* in *cis* to the *lof* allele (*lfr1*, *lfr2* and *WL7*) that were essential for expression of the long-finned phenotype. In a similar approach, γ-ray irradiation of lof^{dt2} founders were used to generate deletions within the *lof* chromosome to map the region of mutation.

In the absence of coding variants within *kcnh2a* in the original mutant line, we sought to identify the causative genetic lesions underlying the *lof* phenotype. Given the suppression of recombination over a large mapping interval (~1 Mb), and to account for the possibility of structural variants in *lof*, we reconstructed the chromosome 2 locus using long sequencing reads from a lof^{dt2} incross (NCBI Sequence Read Archive ERX1428166). As the revertant screens identified *kcnh2a* as necessary for the *lof* phenotype, we searched for single molecule PacBio sequencing reads from a lof^{dt2} incross that contained a portion of *kcnh2a* using BLASTN (ncbi-blast-2.2.30+; parameters '-max_target_seqs 1 -outfmt 6'). Candidate reads were then assembled into contigs using the Canu assembler[65] (parameters 'genomeSize = 0.01 m -pacbio - -minInputCoverage = 2 stopOnLowCoverage = 2′). Both individual reads and the assembled contigs identified a breakpoint upstream of the transcription start site of *prrx1a* (~23.174 Mb in Zv11) and one between *kcc4a* and *kcnh2a* (~24.075 Mb) (Figure S3B). We further refined these contigs by identifying reads that contained a high-quality match (BLASTN) to each side of the identified breakpoint. This merged assembly resulted in two contigs, one for each breakpoint. Notably, we did not generate contigs representing the wild-type (Zv11) chromosome organization at these breakpoints. The breakpoint between *kcc4a* and *kcnh2a* is within a DNA transposon that is not present between these genes in the zebrafish reference genome (Zv11) and is likely part of a transposable element expansion in this region in *lof* relative to the Tübingen strain (Figure S3C).

Overexpression constructs

pmTol2-ubi:hKCNH2, ubi:eGFP was constructed by inserting a ubi:hKCNH2 and a ubi:eGFP cassette into the multiple cloning site of pminiTol2 (Addgene #31829). The ubi:hKCNH2 and ubi:eGFP cassettes were obtained by inserting either the full length coding sequence of hKCNH2 with the SV40 late polyadenylation signal (SV40pA) or eGFP from pME-eGFP (Tol2kit #383) downstream of the ubiquitin promoter of pENTR5′_ubi (Addgene #27320). Plasmids (20 ng/μl) and Tol2 transposase mRNA (25 ng/μl) were injected into single cell zebrafish embryos. Adults were screened for GFP+ cell clones containing the integrated plasmid and the impact of GFP+ fin clones on fin growth was assessed.

Isolation and mapping of nr21 mutant line

nr21 was isolated in a small-scale dominant screen. In this screen, wild-type male founders were mutagenized with ENU as above and first-generation outcross progeny screened for fin phenotypes. A single *nr21* founder was isolated as having a dominant, shortened length of all fins compared to wild-type. This mutant was outcrossed to establish the mutant line. As *nr21* is dominant and without an obvious homozygous phenotype, we could not easily isolate homozygous *nr21* individuals for mutation mapping by homozygosity by descent.[15] Instead, we used these methods to identify the recessive, wild-type chromosome. However, given normal variability in zebrafish strains, more than one haplotype of wild-type chromosome can be present in the F2 generation thus appear as heterozygous (e.g., AB1, AB2). To reduce the likelihood of mixing parental wild-type chromosomes (AB1/AB2), three wild-type F2 families were mapped to identify crosses containing homogeneous chromosomes (Figure S2D). Next generation sequencing libraries were made from pools of 10 wild-type siblings from each of three separate F2 crosses. Additionally, in order to identify candidate mutations, we also sequenced a pool of 10 short-finned *nr21* individuals from a mixture of all three F2 families. DNA was extracted from fin clips using QIAGEN DNeasy Blood & Tissue Kit. Barcoded sequencing libraries were prepared as in Bowen et al.[71] and hybridized to a custom Agilent SureSelect 1M Capture Array (cat # G3358A) targeting the zebrafish exome (Zv9). The zebrafish design encompassed 974,016 probes of 60 bp length that are tiled on average every 40bp (20bp overlap) over the 41 Mb exome. Sequencing libraries were sequenced as 100bp single end reads on an Illumina HiSeq 4000. Sequencing reads were aligned to the zebrafish genome with Novoalign (http://www.novocraft.com/), and patterns of heterogeneity used to map the *nr21* chromosome as in Bowen et al.[15]

CRISPR-Cas9 gRNA design and injection

CRISPR guide RNAs (gRNAs) were designed against the third coding exon of *lat4a* using the online ChopChop tool to limit predicted off-target gRNA cutting.[63,64] Given the variable success rates of individual gRNAs, a "blanket" of 5 gRNAs targeting the same exon were synthesized and injected as a pool. gRNAs were assembled according to Gangon et al.[41] Briefly, oligos containing the gRNA target were annealing to a universal oligo containing the tracrRNA and SP6 promoter. The annealed oligo ends were then filled in with T4 polymerase for 20 minutes at 12°C. The gRNA was synthesized from this oligo template using Ambion MEGAscript SP6 Kit. For transcription efficiency, the first two bases of each gRNA were changed to 'GG' as there is evidence that these bases have less effect on Cas9 cutting efficiency or off-target binding than mutations closer to the PAM site.[72–74] gRNAs were injected into single cell zebrafish embryos at a concentration of 50 ng/μl gRNA pool and 600 ng/μl Cas9. To screen for deletion efficiency, the target exon was amplified from pools of three 24-hour embryos and the resulting amplicons were heated to 95°C and cooled at −0.1°C per second to form heteroduplexes. Following heteroduplex PCR, a T7 endonuclease digestion for 30 minutes at 37°C in NEB Buffer 2 was used to generate deletions in the presence of Cas9-induced indels. PCR primers for T7 endonuclease assay: 5′-CACTGAAAACTGACCAC AAGTCA-3′, 5′CAGCACTTCCCAGAGTGTCA-3′.

Full gRNA oligo (target sequence underlined, 5′-3′):

1. ATTTAGGTGACACTATAGGCCCTGTACCGTTACCTGGGTTTTAGAGCTAGAAATAGCAAG
2. ATTTAGGTGACACTATAGGTGAATGCCACAAGACTTGGTTTTAGAGCTAGAAATAGCAAG
3. ATTTAGGTGACACTATAGGATGCCACAAGACTTGAGGGTTTTAGAGCTAGAAATAGCAAG
4. ATTTAGGTGACACTATAGGCCACAAGACTTGAGGAGGGTTTTAGAGCTAGAAATAGCAAG
5. ATTTAGGTGACACTATAGGAGTGTGATGGCGCTCAGGGTTTTAGAGCTAGAAATAGCAAG

Universal constant oligo:

5′-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3′

qRT-PCR

Uncut and regenerating caudal fins were homogenized in TRIzol Reagent (Invitrogen) for RNA extraction. cDNA synthesis was performed SuperScript III Reverse Transcription Kit (Invitrogen) and oligo dT primers. We performed qRT-PCR using the Power SYBR Green Master Mix (Applied Biosystems) on the Applied Biosystems ViiA 7 Real Time PCR System. Cycling conditions: 10 minutes at 95 °C; 40 cycles of 15 seconds at 95 °C followed by 1 minute at 58 °C; melting curve analysis with 15 seconds at 95 °C, 1 minute at 58 °C and 15 seconds at 95 °C. Temperature was varied at 1.6 °C/s. In caudal fins, expression levels of *kcnh2a* were normalized relative to *gapdh*. In pectoral and dorsal fins, expression levels of *kcnh2a* were normalized relative to *β-actin.* Fold expression of *kcnh2a* and *lat4a* in wild-type and *lof* and *nr21* backgrounds was calculated using $2^{-(\Delta\Delta Ct)}$. *kcnh2a* primers: 5′-CCTCATCTTCAGTTCCCTCT AAC-3′, 5′-GCTGCCTGTGCATCATTTC-3′. *gapdh* primers: 5′- GGACACAACCAAATCAGGCATA-3′, 5′-CGCCTTCTGCCTTAACC TCA-3′. *β-actin* primers: 5′-CCCAAAGCCAACAGAGAGAA-3′, 5′-ACCAGAAGCGTACAGAGAGA-3′. *lat4a* primers: 5′-ACGCAGG ACGGCAATAAA-3′, 5′-GCTGAAGATACTCCGCATGAA-3′.

Targeted Sequence Capture Design

We based the sequence capture design primarily on the medaka reference genome (*Oryzias latipes*, MEDAKA1), which was the only genome from Beloniformes available when this study was initiated. However, to account for the possibility that specific genetic regions may be absent or highly divergent in the medaka genome but conserved in the suborder Belonoidei, we included regions from the outgroup genomes of *Poecilla formosa* (Poecilia_formosa-5.1.2) and *Xiphophorus maculatus* (Xipmac4.4.2). To identify annotations in *P. formosa* and *X. maculatus* that were absent or not well-conserved in *O. latipes*, we used BLASTN (ncbi-blast-2.2.30+; parameters '-max_target_seqs 1 -outfmt 6'). Any element from *P. formosa* or *X. maculatus* with an E-value > 0.001 and/or covered < 70% in the *O. latipes* genome was included in the capture. If the element missing in medaka was found in both *P. formosa* and *X. maculatus*, we included the *X. maculatus* version. To avoid adding DNA regions that are specific to Poeciliidae, DNA sequences from *P. formosa* and *X. maculatus* were identified by BLASTN (E-value < 0.001) in the genome of an additional outgroup, the Nile tilapia (*Oreochromis niloticus*; Cichlidae ; Orenil1.0), prior to inclusion in the capture. SeqCap EZ Developer (cat #06471684001) capture oligos were designed in collaboration with the Nimblegen design team to reduce probe redundancy, standardize oligo annealing temperature and remove repetitive regions.

We generated two sequence capture designs for targeted enrichment of beloniform sequencing libraries, one for protein-coding exons and another for conserved-non-coding elements (CNEs, miRNA, UCNEs). For each genome, protein coding exons were extracted from Ensembl BioMart.[62] CNEs were defined from the constrained regions in the Ensembl compara 11-way teleost alignment.[75] We removed CNEs < 75bp in length to facilitate space in the capture design. In addition to constrained non-coding elements, miRNA hairpins were extracted from miRbase and ultraconservative elements (UCNEs) from UCNEbase and included in the capture design.[76,77] We padded miRNA hairpins to be > 100bp, and removed CNEs, miRNAs and UCNEs that overlapped coding exons using Bedtools (v2.26.0) intersectBed.[50] We prioritized miRNA and UCNE annotations where these overlapped with CNE annotations.

The final capture design for protein coding exons encompassed a total of 225,182 regions and targeting 38,295,260 total bases. The CNE design encompassed 119,080 regions and targeted 17,983,636 total bases. Importantly, we recovered an average of 61.4% read coverage of targeted regions in the genome of *X. maculatus* and 41.5% coverage in *P. formosa*. Recovery of these targeted regions from outgroup genomes that lack clear homology to the *O. latipes* reference genome indicates conservation in Belonoidei with specific loss in *O. latipes*. That we recovered reads from these genomes highlights the importance of a taxonomically diverse capture probe set in minimizing genome bias in cross-species sequence enrichment (Table S2).

Specimen tissue collection and sequencing library preparation

Paired fin and standard fish lengths were measured from beloniform specimens in the Harvard Museum of Comparative Zoology (Data S1A). Tissues for sequencing were acquired from archived samples in the NRL laboratory and the Royal Ontario Museum (Data S1B). DNA was extracted from multiple individuals of each species using the QIAGEN DNeasy Blood & Tissue Kit (Table S1; Data S1B). Equal quantities of DNA from each individual were pooled for each species prior to library preparation and sequencing. The DNA pools were diluted in a shearing buffer (10 mM Tris, 0.1 mM EDTA, pH 8.3) and were mechanically sheared to an average size of 200bp in a Covaris E220 ultrasonicator (duty cycle, 10%; intensity, 5; cycles/burst, 200; time, 400 seconds; temperature, 8°C). Barcoded sequencing libraries were generated using a KAPA HyperPrep Kit (Roche, No. 07137923001) following the Nimblegen SeqCap EZ Library protocol (Version 4.3) and using dual-SPRI (solid phase reversible immobilization) size selection to

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

generate libraries of 200-450 bp. The libraries were hybridized to the capture baits according to the SeqCap EZ Library protocol. In place of Human Cotl DNA, the SeqCap EZ Developer Reagent was used (cat #06684335001) during hybridization. Additionally, libraries were hybridized to the capture baits and washed at a reduced stringency of 45°C relative to the manufacturer recommended temperature of 47°C in order to allow extra mismatches for cross-species hybridization. Multiple barcoded libraries were then pooled for 100 bp single-end sequencing via Illumina HiSeq 2500.

Reference contig assembly

To generate reference contigs for each species, we followed the Phylomapping *de novo* contig assembly pipeline as described in Daane et al.[9] Briefly, low-quality bases in sequencing reads were masked and sequencing reads de-duplicated with the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Sequencing adaptor sequences were removed using Trimmomatic v.0.36.[58] Sequencing reads were binned by orthology to target regions in the *O. latipes*, *X. maculatus* and *P. formosa* reference genomes using BLASTN and dc-megaBLAST. Binned reads were then assembled into contigs *de novo* using CAP3 and UCLUST.[59,60] Sequencing reads were aligned back to the assembled contigs using NextGenMap to recruit previously unmapped reads to the binned assembly.[49] All reads were re-assembled using CAP3 in a second round of *de novo* assembly. If multiple contigs are assembled for any given element (for example, CNE, exon), the multiple contigs were then merged if they overlapped and had > 95% identity. This results in consensus contig sequence for each target region. See Daane et al.[9] for details.

Identification of orthologs

As in Daane et al.,[9] we paired orthologous sequences between species using a gene tree-species tree reconciliation approach. Contigs were automatically annotated according to the orthologous element within the reference genome that was identified using BLAST and subsequently used to scaffold and refine contig assembly. In the event that multiple contigs are assembled for a given targeted element, the multiple contigs for all species were aligned using MAFFT v7.313 (parameters '-op 10 -ep 10'),[54] and a maximum likelihood tree generated with IQTree v1.7beta2 (parameters '-alrt 1000').[47] To infer patterns of duplication and loss, gene tree reconciliation was performed using Notung-2.9 (parameters '–reconcile–rearrange–silent–threshold 90%–treeoutput nhx').[66] The total number of duplication and loss events across the phylogeny as estimated by Notung was then compared to a scenario where all copies are local duplicates. The most parsimonious scenario where the fewest gain/loss events occurred was then selected. Simulations of this ortholog pairing approach found accurate segregation of orthologs provided there $\geq$ 4%–6% variation between paralogous sequences, which coincides with thresholds of variation necessary to distinguish copy number variation during contig assembly (Data S1E).[9]

Read coverage and depth of targeted regions

The targeted sequence captures were designed from existing reference genomes, which were then used to scaffold contig assembly. In order to assess capture efficiency and estimate the coverage of targeted regions, we need to identify where the coordinates on each assembled contig correspond to on the original reference genomes. We performed a pairwise alignment between each assembled contig and the targeted region of the genome using Biopython v.1.70 (parameters 'pairwise2; match = 5, mismatch = −4, gap_open = −15, gap_extend = −1'). Raw sequencing reads were re-aligned to the assembled contigs using NextGenMap v0.5.5,[49] and the data from the pairwise alignment, including indels relative to the reference genome, were used to lift the read alignment data from the contig to the reference genome. Alignments were converted from SAM to BAM format and indexed using SAMtools (v1.9)[51] and were visually inspected for accuracy in the Integrative Genome Viewer.[52] We defined coverage as the number of targeted bases overlapping at least one sequencing read. Coverage was calculated with BEDTools v2.23.0,[50] with read depth per base calculated with using the depth flag of coverageBed (parameters '-d'). Coverage was estimated at multiple levels of read depth (Table S3).

Among the genes with low coverage, we detected enrichment for gene classes associated with the cell adhesion, extracellular matrix proteins and the immune system (< 25% coverage in $\geq$ 75% of exons; Data S1C). Enrichment was calculated with a Fisher's exact test (SciPy v.0.18.1; fisher_exact), and multiple hypothesis test corrections using Benjamini-Hochberg FDR (python module statsmodels v.0.6.1; fdrcorrection0). The missing genes are comparable to similar cross-species targeted sequence enrichment experiments for other clades,[8,9] indicating that these fast-evolving gene classes are less likely to be highly represented in these datasets.

Nucleotide diversity and recovery of population variation

Where possible, we performed pooled sequencing of multiple individuals for each species (Table S1; Data S1B). We calculated genome-wide nucleotide diversity (π) across all sequenced regions using the program PoPoolation v1.2.2 (parameters '–min-count 2–min-coverage 4').[53] We also distinguished population variation from polymorphisms that are fixed within the species (Data S1D). Mutations in the dataset for each species were considered heterozygous with a minimum allele depth of 2 sequencing reads and an allele frequency between 20%–80% in the sequencing reads.

Multiple sequence alignment

We used the program MAFFT v 7.313 to align orthologous sequences across our dataset (parameters '–maxiterate 1000 –localpair –op 10 –ep').[54] In the event the MAFFT alignment had an indel that disrupted a codon, we re-aligned these exons as codon alignments using the frameshift-aware multiple sequence alignment program MACSE v2.03 (parameters 'prog alignSequences -seq -seq_lr -fs_lr 10 -stop_lr 15').[55] We masked the multiple sequence alignment using the program Spruceup (v 2020.2.19, parameters

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

'data_type:nt, distance_method:uncorrected, window_size:6, overlap:5, fraction:1, criterion:lognorm, cutoffs:0.97').[56] Spruceup uses a lognormal distribution of genetic distances across the multiple sequence alignment to detect local outliers in the alignment. For more accurate estimation of genetic distances, we ran Spruceup on a concatenated set of all contigs in our multiple sequence alignment instead of on the much smaller individual exons or CNEs that make up the contigs.

Beloniformes phylogeny

We used IQTree v1.7beta2 (parameters '-bb 1000 -st CODON -m MFP') to calculate a maximum likelihood tree for each gene $\geq$ 400 bp in our dataset in which there was $\geq$ 70% coverage across all species (4,683 genes).[47] For each gene we used ModelFinder to find the optimum codon model,[78] and assessed support for phylogenetic relationships using 1,000 ultra-fast bootstrap replicates.[79] We then estimated a species tree from the distribution of gene trees using ASTRAL v5.6.1,[57] with local posterior probabilities to support each quadpartition in the tree.[80]

Reconstruction of gene sequences

We reconstructed full gene sequences from our individual exon contig data. As in Daane et al.,[9] we concatenated single copy exons in gene order as found in the reference genomes that were used to guide contig assembly. We used the *Oryzias latipes* HdrR (Japanese medaka; MEDAKA1), *Poecilia formosa* (Amazon molly, Poecilia_formosa-5.1.2), or *Xiphophorus maculatus* (platyfish, Xipmac4.4.2) reference genomes to guide exon concatenation. A total of 23,162 gene sequences were reconstructed for each species. In some analyses, a candidate set of known fin and limb-associated genes were specifically assessed (Data S1H).

CNE association with genes

Many CNEs function as tissue specific enhancers and promoters. As in Daane et al.,[81] we used the Genomic Regions Enrichment of Annotations Tool ('GREAT') approach to assign CNEs to neighboring genes as putative regulatory targets.[30] This approach defines a basal regulatory window that is 5kb downstream and 1kb upstream of each gene's transcription start site (TSS). This window is then extended up to 1Mb upstream and downstream of the TSS or until overlap the basal regulatory window of another gene. As opposed to simple nearest neighbor approaches, this approach allows CNEs to fall within the window of one or multiple genes, enhancing statistical power for gene ontology enrichment studies.[30] We based the regulatory windows in our analysis on the *Oryzias latipes* reference genome.

Ontology annotations

To supplement the gene functional annotations for the Japense medaka, Amazon molly and platyfish orthologs we generated a merged gene ontology dataset from multiple species. We mined gene ontology annotations from human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus*), three-spined stickleback (*Gasterosteus aculeatus*), medaka (*Oryzias latipes HdrR*), and zebrafish (*Danio rerio*) in Ensembl BioMart (downloaded January 2020).[62] All Ensembl gene identifiers were then converted to a non-redundant set of medaka orthologs.

Evolutionary sequence rate analysis

We calculated branch lengths for each gene along a fixed species tree topology using IQTree v1.7beta2 (parameters '-te -keep-ident -st DNA -m MFP').[47] We then used the program RERConverge v0.1.0 to estimate the relative evolutionary rates for each gene at all nodes in the phylogeny (parameters 'transform = ''sqrt,'' weighted = T, scale = T, cutoff = 0').[48]

To track genome-wide patterns in relative evolutionary rate, we averaged the relative evolutionary rate across all genes in a given gene ontology term to generate a gene ontology-wide average relative evolutionary rate for each node in the phylogeny (Data S1F). We then compared the mean relative evolutionary rate between the gliding beloniform nodes and the other beloniform species using a Wilcoxon signed-rank test. P values were corrected using FDR (Python module statsmodels v0.6.1; fdrcorrection0). To assess potential sampling bias and noise within the dataset, we calculated differences in relative evolutionary rate for all GO-terms across 500 random samplings of similar tip and ancestral node distribution as the gliding beloniforms (Figure S1B). Only 8 out of 500 samples showed any statistically significant difference in relative evolutionary rate for a GO-term, and all 8 samples had $\leq$ 4 significant terms (Figure S1B).

We also assessed accelerated or constrained sequence evolution along the ancestral branches of flying fishes and halfbeaks using the program phyloP (PHAST v1.4, parameters '–method LRT–no-prune–features–mode ACC').[45,46] The tree model for phyloP was derived using phyloFit and the species tree (Figure 1C). For CNEs, the tree models were based on 6,874 elements of at least 200 bp and with $\geq$ 75% coverage across all species. For CDS, the phyloP tree models were derived from 2,718 reconstructed gene sequences of at least 750 bp and with $\geq$ 75% coverage across all species (Data S1J). We calculated gene ontology enrichment for accelerated or constrained evolution using a SUMSTAT approach as in Daub et al.[82] (Data S1K–S1N). Briefly, we normalized the log-likelihood ratio test score output from phyloP (ΔlnL) by taking a fourth root (ΔlnL4), and then summed these normalized ΔlnL4 scores across all genes in a given ontology term. P values are estimated by boot-strap resampling (50,000 replicates) and corrected for multiple hypothesis testing using a Benjamini Hochberg false discovery rate procedure (FDR; python module statsmodels v.0.6.1; fdrcorrection0).

Please cite this article in press as: Daane et al., Modulation of bioelectric cues in the evolution of flying fishes, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.054

Analysis of molecular convergence

To search for convergent amino acid substitutions, we looked for amino acid substitutions that were present within both lineages of flying halfbeaks (*Euleptorhamphus viridis* and *Oxyporhamphus micropterus*) and within ≥ 70% of flying fishes (Exocoetidae) but that were not observed in the other Beloniformes (Data S1G). We required coverage of the amino acid position in all immediate sister lineages to the gliding beloniforms (*Rhynchorhamphus georgii, Hemiramphus brasiliensis* and *Hemiramphus far*) and in at least 50% of the other beloniforms in the dataset. To assess background levels of convergent amino acid substitutions in the phylogeny, we calculated the number of identical amino acid substitutions observed in three topologically similar species groupings (Figure S1C). To account for coverage differences among species, we further normalized amino acid counts by the number of bases analyzed and by the total number of SNPs (convergent or not) that are unique to one or more of the foreground species and not found in the outgroups (Figure S1C). This SNP normalization helps control for differences in topology, whereby the relatedness of species may impact the likelihood of having unique SNPs not seen in outgroup lineages.

We assessed convergent evolutionary rates in gliding beloniforms using RERConverge v 0.1.0 with topology weighted correlations to treat each clade as a single observation during p value calculation (parameters for getAllResiduals 'transform = ''sqrt,'' weighted = T, scale = T, cutoff = 0'; parameters for foreground2Tree 'clade = ''all,'' weighted = TRUE'; parameters for correlateWithBinaryPhenotype 'min.sp = 10, min.pos = 5, weighted = ''auto'''; Data S1I).[48] P values were corrected for multiple comparisons within RERConverge using the Benjamini-Hochberg procedure.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of statistical analysis are provided within the relevant figure legends. All statistics were performed using R v3.6.1 (https://www.r-project.org/). Statistical significance of difference between means was assessed using the Wilcoxon signed-rank test (Figure 2A) and two-tailed Student's t tests (Figure 2C). The statistical association between genotype and fin size was performed using one-way ANOVA, with all pairwise comparisons subsequently analyzed using Tukey Honest Significant Difference (TukeyHSD) (Figures 3B, 3D, 4B–4E, S2B, and S2I). Unless otherwise indicated, n refers to the number of individual fish used in the experiment. Multiple hypothesis test correction was performed according to the Benjamini-Hochberg FDR procedure (python module statsmodels v.0.6.1; fdrcorrection0). In data represented by boxplots, the box indicates 2nd and 3rd quartiles, the middle line represents median, and the whiskers denote maximum and minimum values exclusive of outliers. For qRT-PCR data, midline represents the mean and the error bars represent ± 1 standard error of the mean (s.e.m). The dots plotted within figures represent individual data points.