



# Genome resequencing clarifies phylogeny and reveals patterns of selection in the toxicogenomics model *Pimephales promelas*

Katy E. Klymus<sup>1</sup>, Robert A. Hrabik<sup>2</sup>, Nathan L. Thompson<sup>1</sup> and Robert S. Cornman<sup>3</sup>

<sup>1</sup>U.S. Geological Survey, Columbia Ecological Research Center, Columbia, MO, USA

<sup>2</sup>Missouri Department of Conservation, Perryville, MO, USA

<sup>3</sup>U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO, USA

## ABSTRACT

**Background.** The fathead minnow (*Pimephales promelas*) is a model species for toxicological research. A high-quality genome reference sequence is available, and genomic methods are increasingly used in toxicological studies of the species. However, phylogenetic relationships within the genus remain incompletely known and little population-genomic data are available for fathead minnow despite the potential effects of genetic background on toxicological responses. On the other hand, a wealth of extant samples is stored in museum collections that in principle allow fine-scale analysis of contemporary and historical genetic variation.

**Methods.** Here we use short-read shotgun resequencing to investigate sequence variation among and within *Pimephales* species. At the genus level, our objectives were to resolve phylogenetic relationships and identify genes with signatures of positive diversifying selection. At the species level, our objective was to evaluate the utility of archived-sample resequencing for detecting selective sweeps within fathead minnow, applied to a population introduced to the San Juan River of the southwestern United States sometime prior to 1950.

**Results.** We recovered well-supported but discordant phylogenetic topologies for nuclear and mitochondrial sequences that we hypothesize arose from mitochondrial transfer among species. The nuclear tree supported bluntnose minnow (*P. notatus*) as sister to fathead minnow, with the slim minnow (*P. tenellus*) and bullhead minnow (*P. vigilax*) more closely related to each other. Using multiple methods, we identified 11 genes that have diversified under positive selection within the genus. Within the San Juan River population, we identified selective-sweep regions overlapping several sets of related genes, including both genes that encode the giant sarcomere protein titin and the two genes encoding the MTORC1 complex, a key metabolic regulator. We also observed elevated polymorphism and reduced differentiation among populations ( $F_{ST}$ ) in genomic regions containing certain immune-gene clusters, similar to what has been reported in other taxa. Collectively, our data clarify evolutionary relationships and selective pressures within the genus and establish museum archives as a fruitful resource for characterizing genomic variation. We anticipate that large-scale resequencing will enable the detection of genetic variants associated with environmental toxicants such

Submitted 4 April 2022  
Accepted 5 August 2022  
Published 25 August 2022

Corresponding author  
Robert S. Cornman,  
rcornman@usgs.gov

Academic editor  
Khor Waiho

Additional Information and  
Declarations can be found on  
page 22

DOI 10.7717/peerj.13954

Distributed under  
Creative Commons Public  
Domain Dedication

OPEN ACCESS

as heavy metals, high salinity, estrogens, and agrichemicals, which could be exploited as efficient biomarkers of exposure in natural populations.

**Subjects** Aquaculture, Fisheries and Fish Science, Evolutionary Studies, Genomics, Molecular Biology

**Keywords** Fathead minnow, Pimephales, Genome resequencing, Natural selection, Taxonomy, Titin, MTORC1

## INTRODUCTION

Fathead minnow (*Pimephales promelas*) is a key model of vertebrate toxicology in aquatic environments, and standardized tests of acute and chronic toxicity in larvae and adults have been developed to support regulatory science (e.g., [Benoit, Puglisi & Olson, 1982](#); [Pickering, 1988](#)). The reference genome has recently been updated to further support the integration of transcriptomic, proteomic, and epigenetic methods in fathead minnow toxicology ([Martinson et al., 2021](#)). These approaches can help improve the sensitivity and accuracy of toxicological screens, as well as our understanding of molecular modes of action ([Ankley & Villeneuve, 2006](#)).

This wealth of toxicology data contrasts with a dearth of population- and evolutionary-genomic study of the fathead minnow and its relatives. Yet natural genetic variation and evolutionary history can inform toxicology in several ways, as highlighted by extensive work in killifish (genus *Fundulus*). For example, because genetic background can be an important modulator of many traits, including toxicity ([Whitehead et al., 2010](#); [Whitehead et al., 2017](#); [Crawford et al., 2020](#)), this potential source of variation would be important to consider in toxicological assessments ([Baird & Barata, 1998](#)). Genetic-background effects may be particularly strong in inbred research strains (e.g., [Brown et al., 2012](#)), which can experience both strong genetic drift and directional selection imposed by culture environment and passaging scheme ([Latter & Mulley, 1995](#); [Hoffmann et al., 2001](#)). Although there are critical advantages of inbred stocks for phenotypic analysis ([Baird & Barata, 1998](#)) and genetic mapping ([Jansen, 2004](#)), and genetic-background effects are expected to be smaller for larval phenotypes than for adult phenotypes ([Cheverud, Rutledge & Atchley, 1983](#)), the challenge of identifying late-acting, sublethal effects of chronic exposure and relating them to heterogeneous natural populations remains ([Denny, 1987](#); [Horzmann & Freeman, 2018](#)). The broader relevance of candidate genes implicated in toxicological research ultimately depends on the evolutionary stability of those genes within functional networks. The degree to which this holds across taxa is best assessed with comparative genomic approaches such as ortholog clustering, comparative transcriptomics, and synteny mapping ([Wei et al., 2002](#)). Comparative approaches may be particularly important for cyprinid fishes because polyploidization events can complicate the inference of orthology and functional equivalence ([Catchen, Braasch & Postlethwait, 2011](#); [Chen et al., 2019](#)).

Apart from these considerations of how to extrapolate toxicological findings to natural populations and other species, population-genomic variation provides an alternative framework for assessing toxicological effects, one that is increasingly feasible as sequencing

technology continues to improve. This complementary approach seeks to identify spatial and temporal patterns of genetic variation statistically associated with the presence or concentration of toxicants in sampled environments. Such findings provide indirect evidence of toxicant bioavailability and organismal effect that can be further validated in the laboratory. For example, strong environmental clines are often associated with genetic structure at functionally relevant loci (e.g., [Reid et al., 2016](#); [Weetman et al., 2018](#)), and allow additional loci or pathways to be discovered (e.g., [Ferrero-Serrano & Assmann, 2019](#)). These landscape-level associations between toxicants and genotypes may be particularly useful when captive populations fail to predict short- or long-term impacts in the wild, for example when behavioral mechanisms allow populations to avoid high concentrations of a pollutant ([Larrick et al., 1978](#)). Importantly, a high-resolution spatial and temporal record of population-genetic variation in fathead minnow and its congeners can potentially be extracted from the many thousands of specimens currently residing in museum collections (e.g., <http://www.fishnet2.net/>) Collections often span decades and document both the native range of the species and its many introductions elsewhere.

Here we report an initial exploration of genetic variation within *P. promelas* and among its congeners. We particularly sought to evaluate low-coverage resequencing of ethanol-preserved specimens as a means of identifying adaptive genetic variation, applied to a single test population introduced to the San Juan River sometime prior to 1950 ([Nico, Fuller & Neilson, 2017](#)). We also resequenced two samples of each *Pimephales* congener at higher coverage and performed phylogenetic and evolutionary-rate analyses. These efforts support fathead minnow toxicogenomics by clarifying evolutionary histories and patterns, as well as illustrating the utility of resequencing for the discovery of adaptive alleles in this species.

## MATERIALS & METHODS

### Samples

The Museum of Southwestern Biology (Albuquerque, New Mexico) provided fin clips or whole fish from 48 individual specimens for destructive sampling (invoice MSB 20-003), of which 24 were consumed for this pilot study. Sample metadata are provided in [File S1](#) and sample locations are mapped in [File S2](#). Sequenced individuals were originally collected from the San Juan River in the southwestern United States between 2009 and 2014 and had been stored in 95% ethanol. We viewed these samples as favorable for testing DNA extraction and sequencing because of their relatively short time in storage and explicit ethanol concentration. Additionally, the population was established within the past century ([Nico, Fuller & Neilson, 2017](#)), which we reasoned was sufficient elapsed time for adaptation to a novel environment to have occurred (fathead minnow has a generation time of 1–2 years; [Ross, 2001](#)) yet recent enough for selection signatures to still be evident.

### DNA extraction and sequencing

Genomic DNA was extracted from fin clips with the QuickGene Auto12S Nucleic Acid Extraction System and DNA tissue kit (AutoGen, Inc., Holliston, MA, USA) following the manufacturer's protocol. Sequencing libraries compatible with the Illumina HiSeq platform

were generated with the Lotus DNA Library Prep kit (IDT) following manufacturer's instructions. The concentration and fragment-size distribution of libraries were measured with an Agilent TapeStation. Pooled libraries were shipped to Quick Biology (Pasadena, CA, USA) for a single lane of HiSeq X sequencing, with a target coverage of approximately 4X per sample. Although samples were individually indexed to allow variation in sequencing output to be assessed, this level of coverage does not allow robust genotyping of individual samples and therefore analyses of genetic variation in this population were based on pooled data or the consensus haploid genome sequence. Demultiplexed reads assigned to individual samples were deposited in NCBI BioProject [PRJNA807296](#), whereas the pooled reads (which include reads not successfully demultiplexed) were deposited in NCBI BioProject [PRJNA806159](#). The latter data are termed the "SJR pool" hereafter.

For congener resequencing, we used samples previously collected and identified or made new collections as needed (see [File S1](#) for sample details). New collections were taken under State of Missouri Wildlife Collector's Permit (#19224) issued to R. Hrabik, with no additional review required under the relevant animal care policies (see article Declarations for details). Samples were either stored in ethanol or frozen prior to use, and were shipped cold to Admera Health BioPharma Services (South Plainfield, NJ, USA) for DNA extraction and library preparation using the KAPA Hyper Prep kit. Two lanes of 150-bp paired-end sequencing were produced on the HiSeq X platform, multiplexed with three samples per lane, for a target coverage of approximately 20X per sample. Raw sequence data for *Pimephales* congeners has been deposited in NCBI under BioProject [PRJNA806850](#).

For comparison with the SJR pool, we also downloaded short-insert Illumina data ([SRR1304883](#), release date June 10, 2014) upon which the original reference genome was based (FHM 1.0; ([Burns et al., 2016](#))). This was one of two technical-replicate accessions available for BioSample [SAMN02418978](#), an inbred pool reared at a DuPont facility. As the FHM 1.0 assembly was generated primarily by the US Environmental Protection Agency (EPA), we refer to these data as the "EPA pool" hereafter.

### Read processing and mapping to reference genome

For the museum specimens, we experienced a relatively high rate of index error in the form of unexpected index combinations as well as index reads that were shifted by one base, limiting the number of demultiplexed reads assigned to samples. For analyses of factors affecting library success, we assigned read pairs to a sample if the forward and reverse index sequences differed by at most one edit distance each from an expected combination. However, for population-genetic analysis, we mapped all reads as a single pool so that reads unattributed to a specific individual could still be included in the population pool.

All read pairs were trimmed with the bbmap package (<https://gi.doe.gov/data-and-tools/bbtools/>), specifying a minimum quality of 10 and kmer values for matching exogenous adapters of 15 (internal), 11 (3' terminal), and 9 (5' terminal). Reads less than 140 nt after trimming were discarded. FastQC (<https://github.com/s-andrews/FastQC>) was used to evaluate sequence quality and composition before and after trimming.

The updated and annotated reference genome, EPA\_FHM\_2.0 (*Martinson et al., 2021*), was made available by staff of the EPA prior to submission to NCBI. While a parallel gene annotation has since been computed by NCBI using the “GNOMON” pipeline ([https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/gnomon/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/gnomon/)) and official scaffold accessions generated (see accession [GCF\\_016745375.1](https://www.ncbi.nlm.nih.gov/nuccore/GCF_016745375.1)), we continue to use the original EPA scaffold names (which are searchable at NCBI Entrez: <https://www.ncbi.nlm.nih.gov>) and EPA annotation coordinates for predicted protein sequences.

To analyze factors affecting library quality of museum-archived samples, reads were mapped in pairs with bowtie2 v.2.3.4.2 (*Langmead & Salzberg, 2012*) using the “end-to-end” and “very-sensitive” parameter switches. Likely PCR duplicates were marked with Picard (<https://broadinstitute.github.io/picard/>), and read-mapping statistics were computed with the flagstat and stats commands of samtools v. 1.12 (*Li et al., 2009*).

For population-genetic analysis, all congener samples and the two pools were mapped using bowtie2 in paired mode as already described. A mapping quality filter of 30 was imposed prior to creating multisample “pileup” alignments with the bcftools v. 1.12 (*Li et al., 2009*) mpileup command, using the recommended map-quality adjust parameter of 50. Biallelic variants were called using bcftools call, and sites with a quality less than the programmatic maximum ( $Q = 999$ ) were excluded. Called indels and single-nucleotide polymorphisms (SNPs) within three positions of a called indel were not used in any analysis. Consensus sequences were generated for each sample with the bcftools consensus command using default parameters, in which called heterozygous sites are resolved in favor of the original reference sequence.

### Mitogenome assembly and phylogeny

To generate a near full-length mitochondrial alignment, we initially mapped reads to the following mitogenome reference accessions: [AP012102.1](https://www.ncbi.nlm.nih.gov/nuccore/AP012102.1) for *P. vigilax*, [AP012101.1](https://www.ncbi.nlm.nih.gov/nuccore/AP012101.1) for *P. notatus*, and [MG570452.1](https://www.ncbi.nlm.nih.gov/nuccore/MG570452.1) (*Schroeter et al., 2020*) for *P. promelas*. Read mapping was performed with Bowtie2 in paired format with a mapping quality filter of 30, and consensus generation was performed as described above but with a ploidy of one specified. As no reference mitogenome was available for *P. tenellus*, we performed a *de novo* assembly with Spades 3.15.2 (*Bankevich et al., 2012*), specifying a minimum contig coverage of three and skipping the read-correction step due to the high expected coverage of mitochondrial sequences. Approximately complete mitochondrial genomes were recovered for both *P. tenellus* specimens (accessions [OM718773](https://www.ncbi.nlm.nih.gov/nuccore/OM718773) and [OM718774](https://www.ncbi.nlm.nih.gov/nuccore/OM718774)). We elected to repeat this procedure for *P. notatus* as well (accessions [OM718772](https://www.ncbi.nlm.nih.gov/nuccore/OM718772) and [OM743017](https://www.ncbi.nlm.nih.gov/nuccore/OM743017)), because the two samples had a mismatch rate of over 3% when mapped to GenBank accession [NC\\_033941](https://www.ncbi.nlm.nih.gov/nuccore/NC_033941). Assembled mitogenomes were annotated with the MitoFish server (*Iwasaki et al., 2013*).

San Juan River samples were individually confirmed to be *P. promelas* using this consensus-sequence approach, although four samples lacked sufficient coverage for this assessment. Of these 20 consensus sequences, 18 were invariant whereas two had unique and relatively divergent haplotypes indicating multiple introductions. Nonetheless, all 20 consensus sequences had >99% identity to *P. promelas* accessions in GenBank.

Reference mitogenomes for the minnow family Leucisidae were downloaded from NCBI (download date 06/04/21) and appended with the new mitochondrial sequences we had generated. These were then aligned at the nucleotide level with mafft v.7480 (Kato et al., 2002). We partitioned the alignment five ways: (1) all predicted paired rRNA positions, (2) all predicted unpaired rRNA positions, (3) first codon positions, (4) second codon positions, and (5) third codon positions. All other positions, including tRNAs, noncoding positions and a few dual-frame coding positions were excluded. Coding-sequence coordinates were obtained by submitting accession AP011273.1 to the MITOS annotation server (Bernt et al., 2013), whereas rRNA fold predictions were obtained using the RNAWebSuite server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>). A majority-rule consensus phylogeny was produced with MrBayes v.3 (Ronquist & Huelsenbeck, 2003) specifying the general time-reversible substitution model and the “invgamma” model of rate variation. The input and output files of the MrBayes analysis are provided in File S3. To evaluate the consistency of the topology recovered by the Bayesian approach, we also generated maximum-parsimony and maximum-likelihood trees in MegaX v. 10.0.5 (Kumar et al., 2018) using the same alignment. The maximum likelihood analysis also used the general time-reversible model with a gamma + invariant distribution of rate heterogeneity, with five rate classes estimated by the software rather than specified by the user.

### Evolutionary rate analysis

Annotated coding sequence was extracted from each consensus sequence with *gffread* v. 0.12.7 (Pertea & Pertea, 2020) using the longest (or else first-listed) transcript of each gene. Note that codons with unidentified bases (N's) are ignored in evolutionary rate analysis. We first used the PAML package (v. 4.9j; Yang, 2007) to test for positive selection by fitting alignments to two models with codeml: one that specifies two site-specific evolutionary rates (an estimated  $\omega_0 < 1$  and a fixed  $\omega_1 = 1$ ) and one that allows a third evolutionary rate  $\omega_2$  than can exceed one (positive diversifying selection). The recommended test statistic is twice the difference in log-likelihood of the two models, which is then compared to the  $\chi^2$  distribution with two degrees of freedom (Yang, 2007). Because two consensus sequences were available for each species, we randomly grouped these into two four-species alignments and ran codeml on both, using an unrooted tree topology consistent with our nuclear species tree (see Results). We did not include all eight sequences together in the PAML analyses because initial analyses with intraspecific sequences included yielded implausibly high test statistics, indicating that gene trees with very short branches gave biased parameter and likelihood estimations. We therefore considered genes to have significant excess nonsynonymous substitution only if both four-species alignments had a false discovery rate (FDR)-corrected  $P$ -value  $\leq 0.05$ . We excluded short alignments (less than 100 codons) and also those that had elevated transition ratios (parameter  $\kappa \geq 5$ ), because strong transition bias complicates rate estimation (Yang & Nielsen, 2000). A single alignment with a very long tree length was excluded as well, as this implies excessive substitutions or a read-mapping artifact. Genes identified as undergoing positive selection by PAML were then further tested using the related BUSTED method (Murrell

*et al.*, 2015) of the HyPhy package v. 2.5.32 (Kosakovsky-Pond & Muse, 2005). We used all eight sequences in BUSTED analyses because the method does not impose branch and site-specific rates but instead uses a random-effects model to determine the overall likelihood of an unlocalized episode of positive selection. That likelihood statistic is also compared to a  $\chi^2$  distribution with two degrees of freedom (Murrell *et al.*, 2015).

To further validate genes identified by these two methods to be under positive selection in *Pimephales*, we assessed whether positive selection was still supported with the inclusion of additional related cyprinids in each alignment. These expanded alignments included inferred orthologs from zebrafish (*Danio rerio*, Howe *et al.*, 2013), *Cyprinus carpio* (Xu *et al.*, 2014), *Carassius auratus* (Chen *et al.*, 2019), *Sinocyclocheilus rhinoceros* (Yang *et al.*, 2016), and *Puntigrus tetrazona* (synonymous with *Puntius tetrazona*; assembly accession GCF\_018831695). These species were chosen based on their high ranking in BLASTP (Camacho *et al.*, 2009) search results of *P. promelas* predicted proteins as well as the availability of RefSeq models predicted computationally from relatively complete reference genomes. The top protein RefSeq match for each of these species was assumed to be orthologous and the corresponding RefSeq mRNA extracted and aligned, deleting poorly aligned regions if necessary. If the expanded BUSTED test was not significant, we also ran the branch-specific method aBSREL (Smith *et al.*, 2015) to confirm that at least one *Pimephales* branch gave evidence of positive selection.

Enrichment of gene ontology (GO) annotations in gene sets was tested with goSeq v. 4.1.1 (Young *et al.*, 2010), which is designed to incorporate length bias in gene sets. Length bias was a potential concern because the power to estimate evolutionary rate classes should generally increase with coding sequence length, which could necessitate a weighting function; however, no relationship was evident. We therefore performed the enrichment test using the standard hypergeometric method that assumes no length bias. To avoid repeated testing of very similar annotation terms, only “cellular compartment” GO categories were used because this category had the lowest *P*-value for an individual term. *P*-values were adjusted using the Benjamini–Hochberg method with the R function *p.adjust* (R Core Team, 2018). The mapping of genes to ontologies was taken from the EPA-submitted annotations (Martinson *et al.*, 2021) by parsing the gff-formatted feature file.

### **Variant calling, genetic diversity statistics, and outlier detection**

To measure variation within the SJR pool, the average per-site minor allele frequency (MAF) was computed directly from the allele counts reported in the mpileup file, *i.e.*, without regard to whether a site was called a variant or not in the multispecies comparison. For these estimates, we ignored sites with fewer than 15 aligned reads or with only a single variant read. MAF was summed across all nonmajor alleles if more than two alleles were reported at a site. This approach was preferred for this SJR-specific statistic so as not to be limited to sites called from a heterogeneous combination of intraspecific population pools and individual congener samples. However, the called and filtered multisample SNPs were necessarily used to calculate Weir’s  $F_{ST}$  with VCFtools (Danecek *et al.*, 2011), with each species grouped as a population. The EPA pool was excluded from the  $F_{ST}$  calculation

because it was strongly inbred and thus not indicative of natural variation. Alignment depth was calculated in sliding windows with VCFtools and for bed-formatted exons with samtools.

Statistical packages are available for identifying sliding-window variation in allele frequencies and  $F_{ST}$  between two population pools (e.g., [Kofler et al., 2011](#); [Kofler, Pandey & Schlötterer, 2011](#); [Boitard et al., 2013](#)). However such differences would not be meaningful for understanding adaptation in the SJR pool given the strong but mosaic inbreeding in the EPA pool (see Results). We therefore performed empirical assessments of outlier polymorphism patterns within the SJR pool only, which for individually-genotyped samples is often implemented by defining outlier quantiles of Tajima's D statistic and nucleotide diversity ( $\pi$ ) ([Feulner et al., 2013](#); [Reid et al., 2017](#)). However, estimation of these statistics from high-throughput sequencing can be fraught, particularly for pooled data with unequal representation of relatively few chromosomes ([Tajima, 1983](#); [Futschik & Schlötterer, 2010](#); [Ferretti, Ramos-Onsins & Pérez-Enciso, 2013](#); [Boitard et al., 2013](#)). Interpretation of Tajima's D is further complicated by demographic factors ([Ramírez-Soriano et al., 2008](#)) such as the likely bottlenecks and admixture of the SJR population during colonization (see Results). We therefore followed an empirical approach similar to that of [Reid et al. \(2016\)](#), but rather than using separate outlier cutoffs for MAF and  $F_{ST}$ , we sought to identify consecutive genomic windows that were visual outliers in a two-dimensional plot of MAF and  $F_{ST}$ . We adopted this two-variable approach because allele frequencies and  $F_{ST}$  at a locus are generally not independent yet modeling their dependence is challenging ([Flanagan & Jones, 2017](#)). To minimize stochastic effects on allele-frequency estimation, we averaged MAF over large genomic windows (50 kb with a 25-kb step), required a minimum of 100 sites per window for  $F_{ST}$ , and required consecutive windows to be outliers. Our specific procedure was to first identify candidate scaffolds with multiple windows at the margins of the plotted two-variable distribution, and then to plot those variables linearly along those candidate scaffolds for inspection. Contiguous regions with conspicuous outlier windows were then re-plotted relative to the total distribution of all windows to confirm that they were collectively near the margins of the distribution. Although we considered the two-variable distribution to be more useful for identifying outliers, we nonetheless confirmed that all selective-sweep regions identified by that approach included at least one window in the 1% quantile tails for both MAF and  $F_{ST}$  ([File S4](#)). MAF and  $F_{ST}$  estimates for all genomic windows are provided in [File S5](#).

In accordance with US Geological Survey policy for data access, primary data associated with this study have also been released in parallel through an approved repository ([Cornman et al., 2022](#)).

## RESULTS AND DISCUSSION

### Genome coverage from resequencing

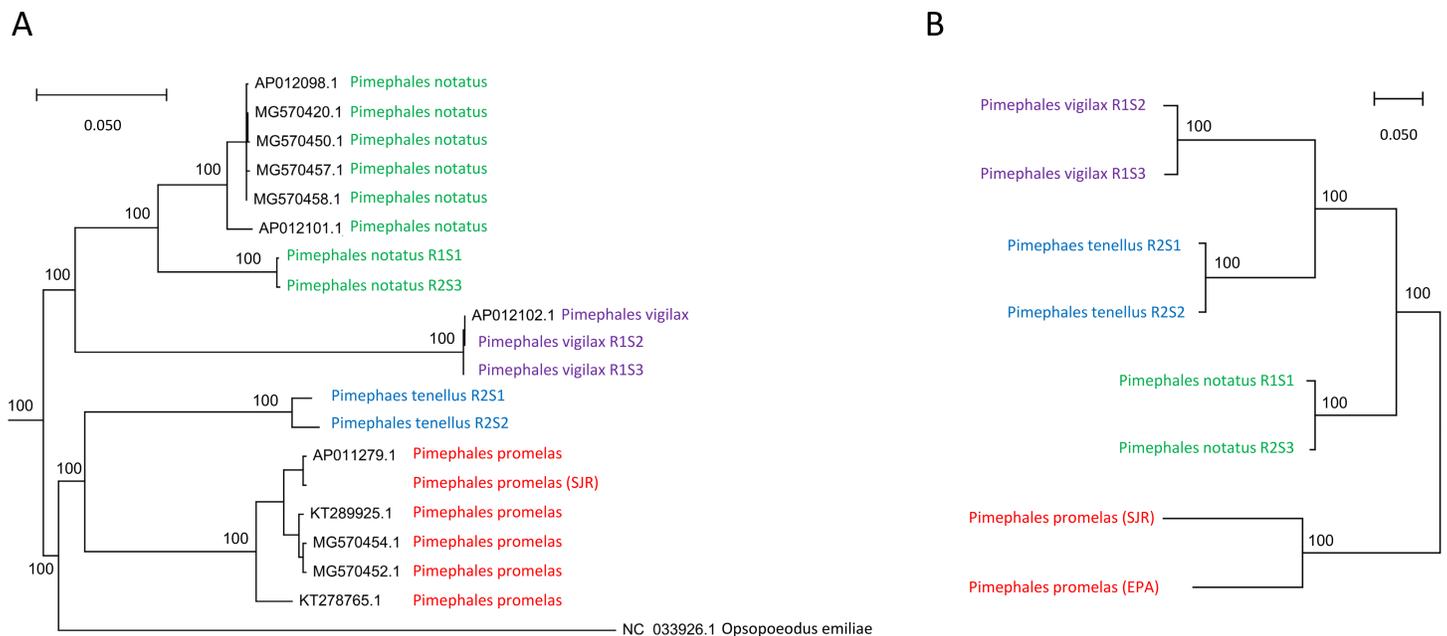
Fin clips from ethanol-stored fish yielded variable but generally adequate DNA quality and quantity ([File S6](#)). Because we do not expect archived samples to yield sequencing libraries of the same quality as fresh samples, we sought to identify library benchmarks

that predicted read-pair mapping and PCR duplication rates. We did not identify any appreciable variation among libraries in rates of PCR duplication, which were modest and averaged 3.9% (File S7). However, we did observe a tendency for lower rates of paired-read mapping to the genome reference when the library DNA concentration was below approximately 3 ng/ $\mu$ l or the peak DNA fragment size was below approximately 450 bp, indicating that these values may be benchmarks for high yield. Peak fragment size was significantly rank-correlated with mapping rate (Spearman's  $r$  of 0.453,  $P = 0.027$ ). We do not suggest that samples yielding libraries below these benchmarks should be avoided, only that such benchmarks can guide the prioritization and relative weighting of samples in resequencing runs. Indeed, we expect that very old or poorly preserved samples will yield relatively fragmented DNA, and also that the relationship between these metrics and sequencing yield will vary among library-preparation protocols. Note that the genome mapping rate is expected to be much less than 100% because the reference sequence is extensively masked, largely due to the high repeat content of the genome (Martinson *et al.*, 2021).

Exon coverage per unit sequencing effort was similar among all congener samples as well as the SJR pool (File S8), averaging 13.8–14.2 read pairs per kilobase per million mapped reads (RPKM) at a mapping quality threshold of 30. Exon coverage in these ranges is expected to be effective for identifying fixed differences among congeners as well as the majority of heterozygous genotypes. Samples from other *Pimephales* species had flatter distributions of coverage and were very similar in shape for the *P. tenellus* pair and the *P. vigilax* pair, whereas the *P. notatus* samples differed somewhat. We attribute the higher frequency of low-coverage exons in congeners to either sequence divergence or paralogy, as both factors reduce mapping quality scores.

### Nuclear and mitochondrial differentiation within *Pimephales*

Eight alignments (six resequenced individuals, the SJR pool, and the EPA pool) were used to ascertain biallelic SNPs. Because the SJR and EPA pools have much higher total coverage and represent a population of chromosomes, SNP ascertainment favors detection of variable sites within those pools. A total of 5,509,510 SNPs were called among the eight samples, of which 4,712,129 were genotyped in at least one congener. The latter SNP set was used to calculate a neighbor-joining phylogeny (Fig. 1) based on average pairwise identity by state (IBS). The SJR and EPA pools were treated as pseudodiploid samples for this analysis so that a pairwise IBS could be calculated (biallelic sites in these pools would typically be represented as heterozygotes). This approach better accommodates intraspecific variation within congener samples, at a cost of distorting the divergence between the two fathead minnow pools. All nodes of the resulting tree had 100% bootstrap support and the topology is consistent with hypothesized relationships based on morphology (reviewed by Schmidt, Dowling & Gold, 1994), specifically that *P. tenellus* and *P. vigilax* are more closely related whereas *P. notatus* is sister to *P. promelas*. Inspection of individual gene trees used in the evolutionary rate analysis (described below), which were based on codon-aware nucleotide substitution models rather than simple differences in state, showed them to be entirely consistent with the IBS tree.



**Figure 1** Phylogenetic analysis of *Pimephales* mitochondrial and nuclear variation. (A) Bayesian phylogeny of full mitochondrial ribosomal RNA (rRNA) and protein-coding sequence. For model estimation, rRNA sites were partitioned into paired and unpaired sites, whereas coding sequence was partitioned by codon position. (B) Neighbor-joining phylogeny of nuclear single-nucleotide polymorphisms (SNPs) using the identity-by-state (IBS) distance measure. The *P. promelas* samples are population pools represented as pseudodiploid samples, *i.e.*, variant sites in those pools are generally represented as heterozygotes resulting in exaggerated intraspecific branch lengths.

Full-size DOI: 10.7717/peerj.13954/fig-1

A mitochondrial alignment of over 15,000 positions across five partitions and including numerous outgroups produced a *Pimephales* topology inconsistent with the nuclear tree (Fig. 1) but consistent with the mitochondrial topology of Bielawski, Brault & Gold (2002). Those authors noted the conflict between this mitochondrial topology and several morphological characters, and suggested that those traits may have evolved independently in parallel. Given the amount of data supporting our nuclear phylogeny, we consider interspecific mitochondrial capture to be a more likely explanation of discordance, possibly from *P. promelas* to *P. tenellus*. Hybridization among cyprinids is common, and mitochondrial capture without apparent impact on nuclear genetic relationships has been reported even between relatively distant species pairs (Freyhof *et al.*, 2005). Figure 1 also shows a relatively deep phylogeographic division within *P. notatus* (*P. notatus* accessions in Fig. 1 originated from Atlantic watersheds, if reported, whereas our *P. notatus* samples were midwestern in origin). Overall, these results are in line with other instances of phylogeographic structure and nuclear-mitochondrial discordance reported for *Pimephales* populations (Ballesteros-Nova *et al.*, 2019; Schönhuth *et al.*, 2016).

The full mitochondrial tree showing the inferred relationship of *Pimephales* to other Leucisidae is shown in File S9. The Bayesian tree placed pugnose minnow (*Opsopoeodus emiliae*) within *Pimephales* with high support, a topology also supported by likelihood and parsimony methods (File S10). Nonetheless, *O. emiliae* represents a long, monotypic evolutionary branch with a novel karyotype (Campos & Hubbs, 1973); its placement could

potentially be influenced by long-branch attraction within the larger cyprinid phylogeny or distorted by reticulate evolution, as we hypothesized above for *Pimephales*. Importantly, [Blanton, Page & Ennis \(2011\)](#) recovered a polytomy for *Pimephales* and *Opsopoeodus* with mitochondrial data but their single-locus nuclear data supported a more distant relationship between the two taxa. Additional nuclear data will likely be needed to place *O. emiliae* with confidence.

The two *P. tenellus* samples derive from different subspecific groups, *P. tenellus tenellus* and *P. tenellus parviceps*, that are morphologically and spatially distinct although intergrades also occur ([Hubbs & Black, 1947](#)). In both the nuclear and mitochondrial trees, the two *P. tenellus* specimens are not more divergent from each other than are any other *Pimephales* conspecific pair. To investigate whether particular genomic windows exist that are highly diverged between the two subspecies, we re-called intraspecific variants for each conspecific pair separately to minimize ascertainment bias with respect to this question. A histogram of the number of fixed differences between sample pairs in 50-kb genomic windows does not indicate any high-divergence windows between *P. tenellus* subspecies ([File S11](#)). We conclude that the two subspecies are not strongly diverged genetically relative to the variation found in comparable sampling of other *Pimephales* species. This result is not inconsistent with strong morphological or ecotypic differentiation, which can accrue from modest allele-frequency shifts across multiple loci, particularly when those loci are in linkage disequilibrium ([Latta, 2004](#)). Although we are not able to evaluate this possibility with our current data, resequencing of *P. tenellus* populations may identify patterns of allele-frequency divergence and linkage disequilibrium that differentiate morphological types.

### Evolutionary rate analysis

We identified 11 genes with evidence of positive selection on coding sequence under both the PAML and BUSTED models (FDR-corrected  $P < 0.05$ ; [Table 1](#)). Most significant alignments were more than 400 codons long and had comparable levels of nucleotide substitution across the species tree ([Table 1](#)). We repeated nine of the eleven tests with expanded alignments that included other cyprinids (see Materials and Methods) to test the consistency of the finding while allowing that positive selection may have also occurred among related taxa (two genes were too divergent in other cyprinids to test). BUSTED found significant evidence of positive selection in eight of the nine expanded alignments, with the remaining gene *nebulin* significant only under the branch-specific method implemented in aBSREL. Increasing the number of taxa, and thus the data available for rate estimation, strengthens the conclusion that these genes have been targets of positive selection. Although the multiple methods used here increase our confidence in these results, important caveats should be noted. First, because genes are not independently annotated in the fathead minnow congeners, we cannot confirm that the true ortholog is actually present in all species or that reads deriving from close paralogs are not also mapping. Second, substitutions are presumed to have fixed along the phylogeny post-divergence, yet retained ancestral polymorphisms and lineage-sorting events may be relatively common among recently diverged species (e.g., [Pollard et al., 2006](#)).

**Table 1** Genes with significant evidence of positive selection within *Pimephales*. *P*-values for *Pimephales*-only alignments are Benjamini-Hochberg corrected for false discovery rate (FDR). A second FDR correction was not imposed on the expanded alignments. *P*-values for *Pimephales*-only alignments are Benjamini-Hochberg corrected for false discovery rate (FDR). A second FDR correction was not imposed on the expanded alignments.

Gene (transcript)	Annotated name	Description	PAML Tree1 <i>P</i> -value	PAML Tree2 <i>P</i> -value	BUSTED <i>P</i> -value	Minimum alignment length	Minimum tree length (substitutions per codon)	Expanded alignment test and <i>P</i> -value**
FMg011803 (FMt022778)	<i>tmc8*</i> ( <i>tmc7</i> )	Transmembrane channel-like protein 7	0.022	0.046	0.001	675	0.087	Not tested
FMg009899 (FMt019254)	<i>plxnc1</i>	Plexin-C1 (semaphorin protein receptor)	0.010	0.001	0.010	1404	0.080	B: 3.675E-7
FMg025023 (FMt045894)	<i>nlrc6</i>	NLR family CARD domain-containing 6	0.008	0.015	0.010	933	0.008	B: 1.21E-3
FMg011037 (FMt021169)	<i>tnfaip2b</i>	Tumor necrosis factor, alpha-induced protein 2b	0.011	0.003	0.025	667	0.102	B: 1.29E-2
FMg006019 (FMt011957)	(unnamed)	Nebulin	0.028	0.011	0.025	6055	0.058	a: 3.22E-2
FMg014331 (FMt027393)	<i>adgre3</i>	Adhesion G protein-coupled receptor E3	0.001	0.000	0.027	612	0.090	B: 2.45E-5
FMg002767 (FMt005485)	<i>alcamb</i>	Activated leukocyte cell adhesion molecule b	0.005	0.030	0.032	558	0.115	B: 0
FMg016591 (FMt031555)	<i>cfbl</i>	Zgc:158446 protein (complement factor b-like)	0.001	0.023	0.032	1200	0.101	B: 1.75E-4
FMg010614 (FMt020395)	<i>scn3b</i>	Sodium channel, voltage-gated, type III, beta	0.040	0.017	0.039	218	0.086	B: 1.24E-3
FMg007724 (FMt015056)	<i>wasf3b</i>	Wiskott-Aldrich syndrome protein family member 3b (WAVE3 homolog)	0.018	0.023	0.039	464	0.058	B: 4.90E-2***
FMg008000 (FMt015520)	uncharacterized	SEA-domain containing	0.012	0.026	0.039	427	0.136	Not tested

**Notes.**

\*Annotated gene name is *tmc8* but is putatively orthologous to *tmc7* of *Danio rerio* based on homology search.

\*\**P*-values computed with BUSTED are marked with "B," whereas *P*-values computed with aBSREL are marked with "a".

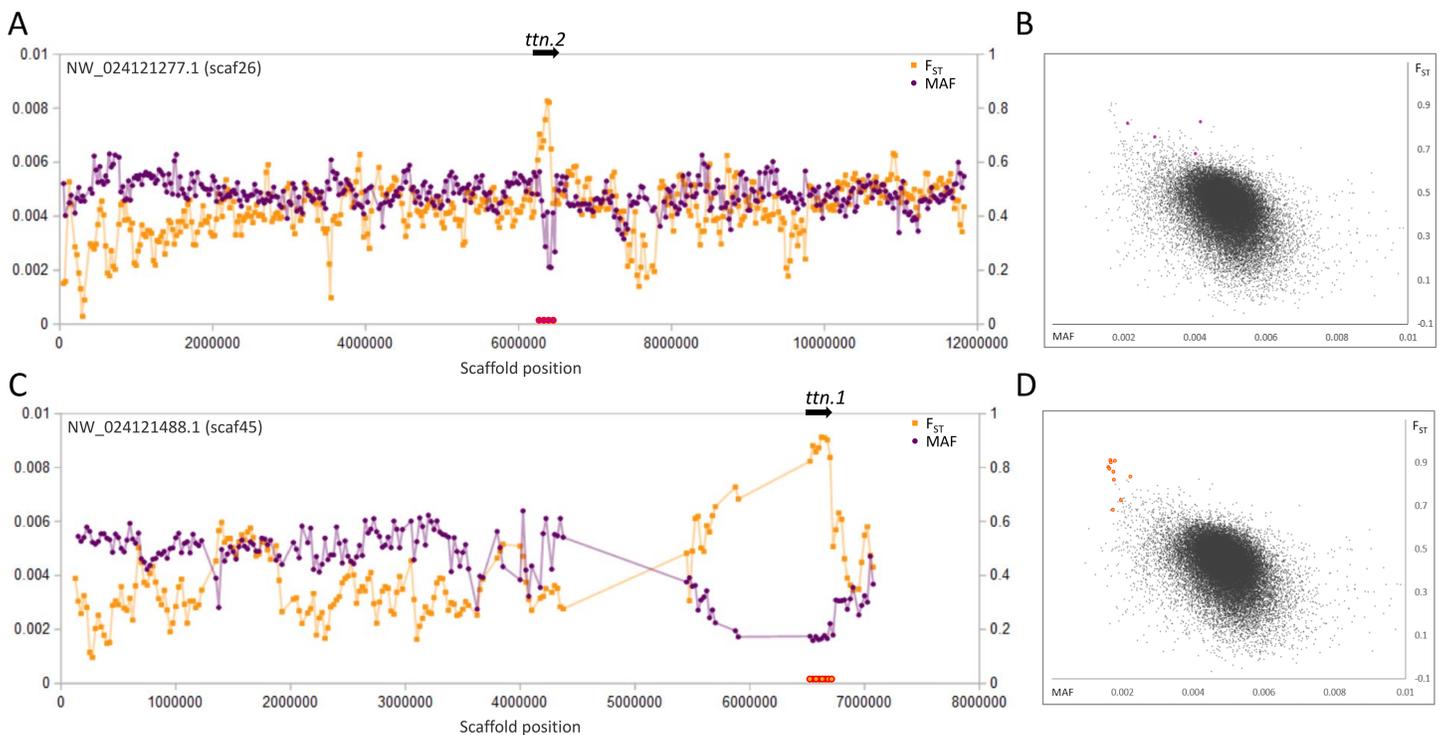
\*\*\**Sinocyclocheilus rhinoceros* removed from expanded alignment due to higher sequence divergence.

At least eight of the eleven positively selected genes encode membrane-associated proteins (Table 1), a high ratio given that their overall proportion in most proteomes is  $\leq 30\%$  (Krogh *et al.*, 2001). Nonetheless, no cellular-component ontology terms were significantly enriched after FDR correction. Among these membrane-associated genes was the sodium transporter *scn3b*, which was also identified as a target of positive selection in the cyprinid *Gymnocypris przewalskii* (Tong & Li, 2020). *G. przewalskii* inhabits high-salinity environments in central Asia, and *scn3b* was one of several solute transporters implicated in salt tolerance in that species. In *G. przewalskii*, *scn3b* is not gill-associated but rather expressed in a variety of tissues at modest levels (Tong & Li, 2020). A second membrane transporter gene, annotated as *tmc8* but most similar by protein sequence to *tmc7* in other vertebrates, is one of a family of well-conserved transmembrane-channel proteins of unknown function. The *tmc7* gene was one of many associated with growth-rate variation in common carp (Su *et al.*, 2018).

Two of the eleven positively selected genes have known immune functions: *cfbl* encodes a complement system component (Zhang & Cui, 2014) and *adgre3* encodes a surface receptor expressed by various myeloid immune cells (Hamann *et al.*, 2016). The *nlr6* gene is a member of the diverse NLR gene family that contributes to pathogen recognition in innate immunity (Hu *et al.*, 2014; Sahoo, 2020). However, the NLR-C subfamily is highly expanded and variable in copy number in fish (Laing *et al.*, 2008), such that it is difficult to ascribe functions to individual gene-family members. Two selected genes have been implicated in axon guidance during development: *alcamb* is important in guiding retinal ganglion cells in *D. rerio* (Diekmann & Stuermer, 2009) whereas the semaphorin-receptor *plxnc1* can function in axon guidance but may also participate in other developmental processes (Jongbloets & Pasterkamp, 2014). The *wasf3b* gene is a paralog of mammalian *wasf3/wave3*, which regulates the cytoskeletal mechanics of cell migration (Kansakar *et al.*, 2020), notably during the development of retinal photoreceptor cells (Corral-Serrano *et al.*, 2020). We speculate that positive selection on both *alcamb* and *wasf3b* reflects adaptation in retinal development, as fish are known to display extreme variation in retinal morphology and spectral tuning (Kusmic & Gualtieri, 2000). The gene *tnfaip2b* is believed to have roles in exocytosis and membrane conformation (Kimura, 2018). The gene *nebulin* encodes a muscle microfiber protein important in muscle structure, contraction, and force generation (Yuen & Ottenheijm, 2020). The final selected gene encodes a predicted protein with multiple domains that are associated with extracellular-matrix glycosylation and cleavage (“SEA” domains, Palmai-Pallag *et al.*, 2005), but is otherwise uncharacterized.

### Selective-sweep candidate regions

We compared polymorphism patterns within the EPA and SJR pools to determine if sliding-window tests of differentiation between the two would be appropriate, knowing that the EPA pool had been inbred prior to sequencing (Burns *et al.*, 2016). This comparison confirmed that the EPA pool was highly inbred overall but also highly mosaic in terms of average MAF within sliding windows, with regions of very low polymorphism intermixed with regions of near normal polymorphism (File S12). Although we cannot determine from a single inbred pool whether this mosaicism is entirely due to chance or also reflects instances of



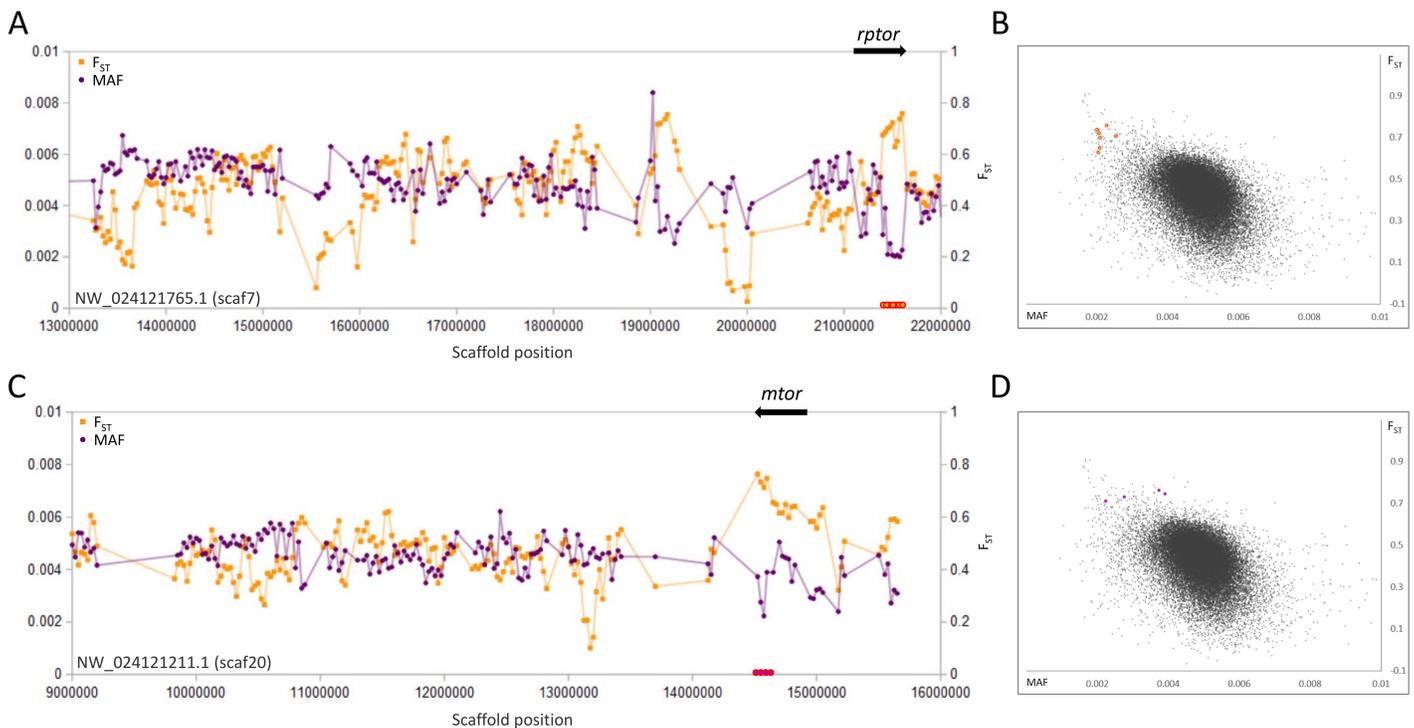
**Figure 2** Genetic variation consistent with selective sweeps in the vicinity of fathead minnow (*Pimephales promelas*) genes *ttn.1* and *ttn.2*. (A) Variation in minor allele frequency (MAF) of the San Juan River pool and among-species  $F_{ST}$  for scaff26, with the location of *ttn.2* marked. Each point represents values within a 50-kb genomic window (sliding, with a step of 25 kb) and is placed at the first coordinate of the window. Absent points indicate too few variants were detected for estimation (see text) and usually reflect repeat-masked regions. Colored circles on the horizontal axis denote windows that appear as outliers in (B). (B) Distribution of MAF and  $F_{ST}$  for all genomic windows analyzed, with the points marked in (A) correspondingly colored. (C) Variation in MAF and  $F_{ST}$  for scaff45, with the location of *ttn.1* marked. (D) Values of genomic windows marked in (C) relative to the distribution of values for all genomic windows.

Full-size DOI: 10.7717/peerj.13954/fig-2

heterosis (higher fitness associated with heterozygosity), resequencing independent inbred lines should distinguish those possibilities, which could be important for understanding sources of genetic variance in toxicogenomic assays.

As the skewed polymorphism in the EPA pool renders it unsuitable as a reference for expected background variation in the SJR pool, we used an empirical approach to identify candidate regions under selection in the latter. We argue that consecutive genomic windows that have high  $F_{ST}$  among congeners and low polymorphism within the SJR pool are consistent with a selective sweep, provided they are sufficiently strong outliers relative to the overall distribution of these values across all genomic windows. Several strong examples of this pattern were found that are both physically clustered and strongly coincident with gene boundaries (Figs. 2–5). Moreover, three cases involved pairs of functionally related genes, providing independent evidence of their biological relevance.

The most pronounced outliers were coincident with the two titin-encoding genes, *ttn.1* and *ttn.2* (Fig. 2), which occur on separate scaffolds in the fathead minnow assembly but are tandemly arrayed in *D. rerio*. Titin is a large, modular protein that helps structure the muscle sarcomere and has spring-like properties, influencing the elasticity of muscle fibers.



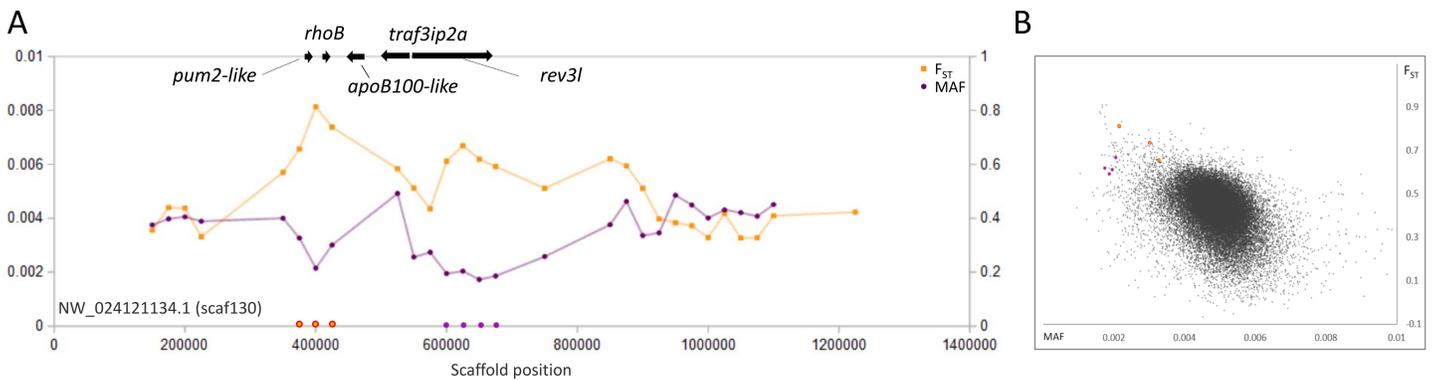
**Figure 3** Genetic variation consistent with selective sweeps in the vicinity of fathead minnow (*Pimephales promelas*) genes *mtor* and *rptor*. (A) Variation in minor allele frequency (MAF) of the San Juan River pool and among-species  $F_{ST}$  for *scaf7*, with the location of *rptor* marked. Each point represents values within a 50-kb genomic window (sliding, with a step of 25 kb) and is placed at the first coordinate of the window. Absent points indicate that too few variants were detected for estimation (see text) and usually reflect repeat-masked regions. Colored circles on the horizontal axis denote windows that appear as outliers in (B). (B) Distribution of MAF and  $F_{ST}$  for all genomic windows analyzed, with the points marked in (A) correspondingly colored. (C) Variation in MAF and  $F_{ST}$  for *scaf20*, with the location of *mtor* marked. (D) Values of genomic windows marked in (C) relative to the distribution of values for all genomic windows.

Full-size DOI: 10.7717/peerj.13954/fig-3

Many different isoforms may be expressed among muscle types and developmental stages (Wang *et al.*, 1991; Steffen *et al.*, 2007).

A second pair of selective sweeps coincided with the genes *rptor* and *mtor* (Fig. 3), the products of which form the key regulatory heteromer MTORC1 (Liu & Sabatini, 2020). These genes lie on different scaffolds and in both cases  $F_{ST}$  is highest and polymorphism lowest at their 3' ends (both genes exceed 100 kb in length and have a concentration of 3' exons). Although other nearby genes cannot be excluded as potential targets of selection, the strong overlap of sweep windows with *rptor* and *mtor*, coupled with their well-documented physical and functional interaction, strongly indicates that they are driving the sweep pattern.

A third example is a pair of outlier regions on scaffold 130, in which the sweep pattern peaks in two 50-kb windows containing the single genes *rhoB* and *rev3l*, respectively (Fig. 4). Although we again cannot rule out the possibility of selection on other nearby genes, functional similarities between *rhoB* and *rev3l* further supports their relevance because both genes function in the repair of genomic DNA lesions. Gene *rev3l* encodes the



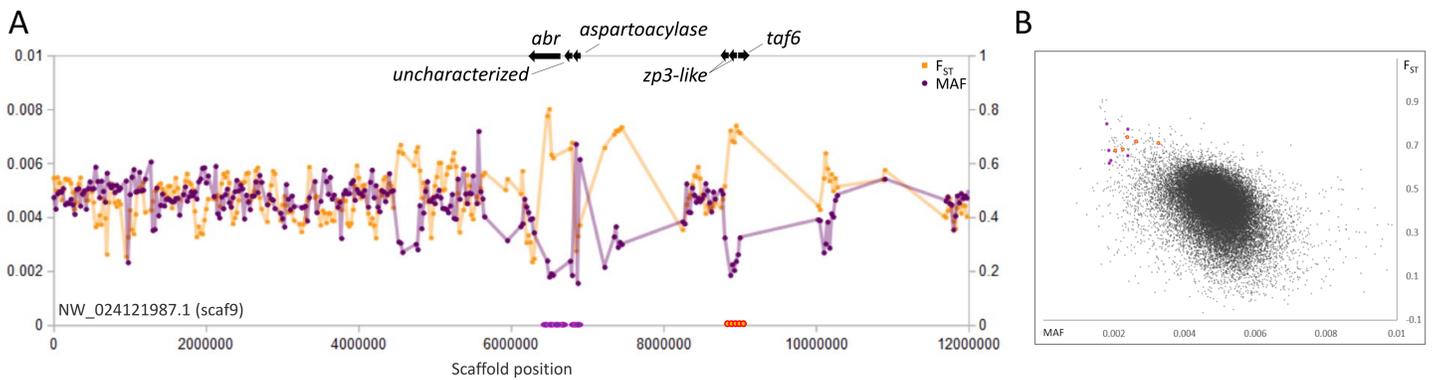
**Figure 4** Genetic variation consistent with selective sweeps in the vicinity of fathead minnow (*Pimephales promelas*) genes *rhoB* and *rev3l*. (A) Variation in minor allele frequency (MAF) of the San Juan River pool and among-species  $F_{ST}$  for scaff130, with the location of genes near outlier windows marked. Each point represents values within a 50-kb genomic window (sliding, with a step of 25 kb) and is placed at the first coordinate of the window. Absent points indicate that too few variants were detected for estimation (see text) and usually reflect repeat-masked regions. Colored circles on the horizontal axis denote windows that appear as outliers in (B). *rhoB* and *rev3l* are the only annotated genes present within the strongest outlier of each set of consecutive windows. (B) Distribution of MAF and  $F_{ST}$  for all genomic windows analyzed, with the points marked in (A) correspondingly colored.

Full-size [DOI: 10.7717/peerj.13954/fig-4](https://doi.org/10.7717/peerj.13954/fig-4)

catalytic subunit of DNA polymerase zeta, which supports translesion replication (Jansen *et al.*, 2009), whereas *rhoB* functions in double-strand-break repair (Mamouni *et al.*, 2014). Both genes can be induced by various types of environmental DNA damage (e.g., Fritz & Kaina, 1997; Lin *et al.*, 2016) and promote the repair of UV-induced lesions (Fritz, Kaina & Aktories, 1995; Jansen *et al.*, 2009). However, they also have regulated expression in normal development (Lange *et al.*, 2016; Vega & Ridley, 2018) and *rev3l* is essential for embryonic viability in mice (Lange *et al.*, 2016). UV exposure in the San Juan River Basin is in fact higher than in the native range of fathead minnow (Fioletov *et al.*, 2004), but whether increased exposure is related to the signatures we identified remains to be tested.

A fourth scaffold, scaff9, contains two distinct sweep regions that do not appear functionally related (Fig. 5). The first region contains the 5' end of the *abr* gene, a general regulator of rho GTPases (Chuang *et al.*, 1995), the uncharacterized gene LOC120472185, and an aspartoacylase-encoding gene (LOC120472146). Aspartoacylase expression is a known transcriptomic biomarker of insecticide exposure in fish (Beggel *et al.*, 2012; Stinson *et al.*, 2022). The second region includes two homologs of zona-pellucida sperm-binding protein 3 (*zp3*) as well as the basal transcription factor *taf6*. ZP3 proteins are documented to evolve rapidly and have been targets of positive selection in diverse taxa including fish (Turner & Hoekstra, 2008; Calkins, El-Hinn & Swanson, 2007; Ugwu *et al.*, 2018).

At present, we are not able to test what environmental factors may be associated with the selective sweeps we observed, as we have not sampled across environmental clines and do not know the progenitor populations. However, titin loci have been implicated in adaptation to environmental variables in other fish species. Whitehead *et al.* (2010) identified distinct titin expression patterns in polychlorinated biphenyl (PCB)-tolerant killifish that may have a cardioprotective role. Milano *et al.* (2014) identified a titin locus as one of several genes containing outlier SNPs consistent with adaptive divergence in European



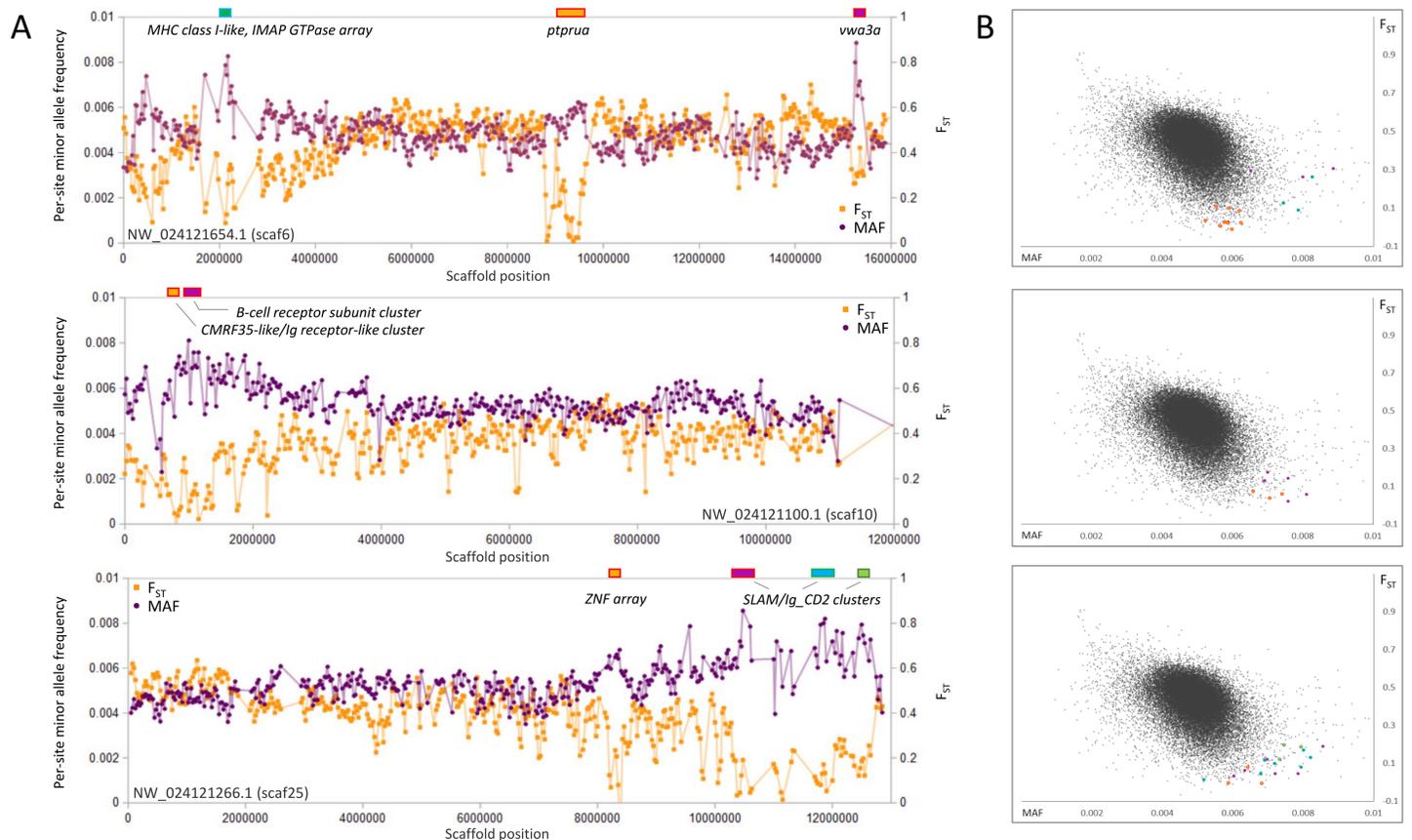
**Figure 5** Genetic variation consistent with selective sweeps on scaffold 9 of the fathead minnow (*Pimephales promelas*) genome assembly. (A) Variation in minor allele frequency (MAF) of the San Juan River pool and among-species  $F_{ST}$ , with the location of genes near outlier windows marked. Each point represents values within a 50-kb genomic window (sliding, with a step of 25 kb) and is placed at the first coordinate of the window. Absent points indicate that too few variants were detected for estimation (see text) and usually reflect repeat-masked regions. Colored circles on the horizontal axis denote windows that appear as outliers in (B). (B) Distribution of MAF and  $F_{ST}$  for all genomic windows analyzed, with the points marked in (A) correspondingly colored.

Full-size DOI: 10.7717/peerj.13954/fig-5

hake, and polymorphism frequencies were correlated with temperature and salinity in some populations. It is also possible that the selective sweeps on titin genes and MTORC1 genes reflect selection on body morphology in novel flow conditions. For example, *Cureton & Broughton (2014)* found repeated evolution of body shape in *P. vigilax* after changes in streamflow, and similar intraspecific associations between body shape and flow regime have been reported in other cyprinids (*Haas, Heins & Blum, 2015; Jacquemin & Pyron, 2016; Kern & Langerhans, 2018*). Adaptive genetic variation associated with the sarcomere proteins titin and nebulin could potentially contribute to such rapid morphological adaptation by altering muscle elasticity. Furthermore, MTORC1 has been shown to be essential for translating experienced mechanical stress into muscle growth (*Goodman, 2013; You et al., 2019*). Alternatively, *mtor* has been hypothesized to act as an evolutionarily labile modulator of how nutrition interacts with growth, development, and body shape (*Valente et al., 2013*), thereby contributing to rapid morphological evolution. This hypothesis has been suggested by several studies implicating *mtor* in ecotype differentiation (*Macqueen et al., 2011*), sexual differentiation (*Manor et al., 2015*), and domestication (*Seiliez et al., 2008; Skiba-Cassy et al., 2009*) in fish.

### Genomic regions of high polymorphism and low $F_{ST}$

In contrast to selective sweeps, consecutive genomic windows with unusually low  $F_{ST}$  among species and high polymorphism within the SJR pool are candidate regions for balancing or frequency-dependent selection (*Fijarczyk & Babik, 2015*). By inspection (*Fig. 6*), clustered outlier genomic windows exhibiting this pattern are often not associated with specific genes but rather arrays of homologous immune genes known to evolve rapidly by both nonsynonymous substitution and copy number. Examples include a cluster of multifamily B-cell receptor subunits (*Clark & Giltiay, 2018*) on scaffold 10 and a cluster of MHC class I (*Shum et al., 2001*) and IMAP GTPase genes (*Huang et al., 2019*) on scaffold 6.



**Figure 6** Genes and gene clusters associated with windows of high intraspecific variation and low among-species  $F_{ST}$ . (A) Variation in minor allele frequency (MAF) of the San Juan River pool and among-species  $F_{ST}$  on three scaffolds, with selected genes or clusters of genes marked. Each point represents values within a 50-kb genomic window (sliding, with a step of 25 kb) and is placed at the first coordinate of the window. Points are absent if fewer than 100 variants were called in the window. Colored boxes on the horizontal axis denote locations of windows that appear as outliers in (B). (B) Distribution of MAF and  $F_{ST}$  for all genomic windows analyzed, with the points marked in (A) correspondingly colored.

Full-size DOI: [10.7717/peerj.13954/fig-6](https://doi.org/10.7717/peerj.13954/fig-6)

Additional immune-gene clusters with similar outlier polymorphism patterns include distinct clusters of SLAM5-like homologs (*Cannons, Tangye & Schwartzberg, 2011*) on scaf25 and a number of CMRF35-like homologs near the B-cell receptor array on scaf10. CMRF35-like genes regulate aspects of inflammation, including eosinophil accumulation in mucosa (*Moshkovits et al., 2014*) and activation of microglial (*Ejarque-Ortiz et al., 2015*) and myeloid cells (*Shik et al., 2014*). All three multigene immune clusters marked in Fig. 6 are associated with a breakdown in synteny relative to zebrafish (File S14), further demonstrating the high rate of structural evolution associated with these regions. Coordinates and descriptions of genes and gene clusters identified in Figs. 2–6 are detailed in File S15.

The overall association that regions of high polymorphism in the SJR pool and low  $F_{ST}$  among species have with immune gene clusters is striking, yet some exceptions are evident in Fig. 6. An array of ten zinc-finger (ZNF) proteins on scaf25 lacks annotated immune functions but does show strong evidence of recent paralogy: six of the ten proteins

cluster together phylogenetically rather than with homologous ZNF proteins identified by BLASTP in the genomes of other cyprinids (File S16). The genes have dissimilar exon structures and several are physically overlapped, indicating that they are not retrogenes or tandem duplicates. A second outlier region on scaff6 coincides with the gene *ptprua*, a conserved single-copy ortholog based on annotation provided on the NCBI gene page (<https://www.ncbi.nlm.nih.gov/gene/563584>) and conserved synteny with the corresponding zebrafish region. The *ptprua* gene encodes a receptor-linked protein tyrosine phosphatase, which as a family have diverse roles in signal transduction, although the PTPRU protein uniquely lacks tyrosine phosphatase activity and instead inhibits the activity of other family members via substrate competition (Hay et al., 2020). A third outlier region on scaff6 is coincident with the *vwa3a* gene, which encodes a cell-surface glycoprotein of unknown function, although the family-defining von Willebrand factor domain is generally found in extracellular matrix proteins and often those involved in cell adhesion (Whittaker & Hynes, 2002). We found MAF to be elevated in two of the three scaff6 regions in the EPA pool as well (File S17), with a strong, narrow spike associated with *vwa3a* but no elevated polymorphism in the vicinity of *ptprua*. This pattern supports the hypothesis of overdominance or frequency-dependent selection at *vwa3a*, manifested even in a captive inbred population.

Although the low  $F_{ST}$  and high diversity in these regions could reflect the retention of relatively ancient haplotypes by overdominance or frequency-dependent selection, as has been shown for immune genes in human (Hughes & Nei, 1989; Slade & McCallum, 1992) and fish (Crisicitiello et al., 2004), the pattern could also be an artifact of paralogy *per se*. That is, the incorrect mapping of reads from recently derived paralogs unrepresented in the reference genome should elevate perceived polymorphism and, if those paralogs are shared among multiple species, reduce perceived  $F_{ST}$ . Haplotype phasing of individuals with long-read sequencing (e.g., Kuleshov et al., 2014) could help disentangle these possibilities. Given the prominence of copy-number variable gene families coincident with these regions, we hypothesize that copy-number variation is the stronger driver of these polymorphism patterns overall, whereas overdominance or frequency-dependent selection more likely explains patterns around the conserved single-copy genes *ptprua* and *vwa3a*. However, a third possible contributor to this polymorphism pattern is sex-associated polymorphism. Sex-associated loci are known to be dispersed in the zebrafish genome, and polymorphisms closely linked to loci involved in sexual differentiation can be maintained at intermediate allele frequencies (Anderson et al., 2012; Wilson et al., 2014). This phenomenon may account for the outlier polymorphism pattern associated with the ZNF array on scaff25, as a cluster of similar ZNF genes are annotated on zebrafish chromosome 4 between 60 and 61 Mb (Ensembl names *Znf1109*, *Znf1033*, *Znf1080*, *Znf1121*, *Znf1130*, and *Znf1044*) and this genomic window has been found by two separate studies to contain SNPs strongly associated with sex in that species (Anderson et al., 2012; Wilson et al., 2014).

### Limitations and caveats

The biological relevance of the selective sweep regions we identified is supported by functional overlap between physically unlinked sweep regions, and our approach further

validated by the detection of patterns in the vicinity of immune gene clusters similar to what has been reported in killifish (*Reid et al., 2017*) and other species (*Hughes & Nei, 1988; Hughes & Nei, 1989; Slade & McCallum, 1992; Criscitiello et al., 2004; Angata, 2006*). Nonetheless, we acknowledge that our current data are not suitable for explicit statistical tests of allele-frequency divergence or linkage disequilibrium relative to a reference population. The detection of selective sweeps can also be complicated by demographic history, including bottlenecks and introgression, as well as the genetic architecture of selected traits (*Storz, 2005*). Although our detection of divergent mitochondrial haplotypes indicates multiple introductions of fathead minnow to the San Juan River, the population is relatively isolated, no major dispersal barriers are apparent within the sampled watershed, and the species has a high growth rate and short generation time (*Ross, 2001*). We therefore do not expect strong genetic structure at neutral nuclear loci in this basin, but this could be confirmed with diploid genotyping.

Although we imposed rigorous thresholds for per-site coverage and total SNPs per window (see Materials and Methods), we nonetheless confirmed that these candidate sweep regions were not associated with atypical levels of sequence coverage that might indicate a mapping artifact ([File S13](#)). Although the two titin genes appear to have somewhat elevated coverage relative to the mean, this likely reflects the much lower level of gapped or masked sequences in those windows due to their very long coding sequences (masked sequence is typically intergenic or intronic). Inspection of coverage patterns for titin exons relative to intron or flanking sequences did not indicate any duplication of annotated exons, and the predicted gene architectures are similar to their orthologs in zebrafish.

## CONCLUSIONS

In this study, we used shotgun resequencing to assess genetic differentiation across the genome for phylogenetic analysis and to detect signatures of natural selection. Our results clarified some evolutionary relationships within *Pimephales*: we found little nuclear or mitochondrial differentiation between morphologically distinct subspecies of *P. tenellus*, whereas mitochondrial data indicated a deep division within *P. notatus*. However, the overall mitochondrial topology of *Pimephales* conflicted with that of nuclear SNPs, implying past mitochondrial capture and undermining the phylogenetic utility of mitochondrial sequences in the genus. An additional result of our genus-level analysis is the identification of eleven genes with excess nonsynonymous substitution (positive selection). While there is no expectation that these genes should overlap with selective sweep regions in the SJR population, it is noteworthy that they included *nebulin*, the product of which interacts directly with titin in structuring sarcomeres. Sequencing other close relatives would be expected to improve evolutionary rate estimates for these genes and allow the timing of selection to be better localized. Including *O. emiliae* in such an effort would be expected to simultaneously clarify the phylogenetic placement of that species.

SNP variation within and among *Pimephales* species revealed several probable selective sweeps in the San Juan River population. We speculate that the prominence of titin genes and MTORC1 subunits in unlinked sweep regions reflects mechanisms of morphological

and metabolic adaptation to new environments, consistent with a body of other work. We anticipate that an analysis of independent invasions of similar watersheds by fathead minnow would illuminate any parallelisms, as has been done in other species ([Fang et al., 2020](#); [Whitehead et al., 2017](#)), which could inform this hypothesis. Identifying commonalities of and constraints on local adaptation during the invasion of novel habitats may help inform the management of longer-lived and more recently bottlenecked invaders such as carp species in North America.

Detection of probable selective sweeps in the SJR population is an important proof-of-principle for using archived samples to detect adaptive alleles associated with toxicants of concern, as many potential toxicological stressors of natural *Pimephales* populations will likely have arisen over comparable time scales. Future work could seek to leverage museum archives associated with documented spatial and temporal clines in pollutants such as industrial hydrocarbons, heavy metals, environmental estrogens, agrichemicals, and salinity. This landscape approach could be extended to model potential responses of cyprinids to climate change, by examining standing variation associated with environmental variables most responsive to climate ([Gautier, 2015](#)). Once candidate haplotypes are identified that are strongly associated with an environmental variable of interest, quantitative PCR assays can be designed to estimate their frequency relative to control markers. Although frequency estimates can be derived for DNA pooled from archived samples or new collections, we ultimately expect the most efficient approach to quantifying contemporary allele-frequencies will be from environmental DNA samples. Conceptually similar eDNA genotyping assays have already been applied to diverse management needs, such as determining the proportion of conspecific carp derived from a non-native source ([Uchii, Doi & Minamoto, 2016](#)), noninvasive genotyping of cultured oysters ([Holman et al., 2019](#)), and noninvasive haplotyping of whale sharks ([Dugal et al., 2022](#)).

## ACKNOWLEDGEMENTS

This study was made possible by the generous donation of archival fathead minnow tissue from the Museum of Southwestern Biology. Tissue for *Pimephales tenellus tenellus* was made available by Jacob Westhoff. Mark Chowning assisted in collecting *Pimephales tenellus parviceps*. Tissue for *Pimephales vigilax* from Nebraska was provided by Sarah Gaughan of the University of Nebraska, Omaha. Trudi Frost provided laboratory support. We are grateful to reviewers who provided their time and expertise to improve the manuscript. The updated and annotated *P. promelas* genome was provided to us prior to public availability by staff of the US Environmental Protection Agency. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by internal funds of the U.S. Geological Survey. No external agencies funded the work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
U.S. Geological Survey.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Katy E. Klymus conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Robert A. Hrabik performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Nathan L. Thompson performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Robert S. Cornman conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

### Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Sampling was authorized under a State of Missouri Wildlife Collector's Permit (#19224) issued to R. Hrabik, and performed consistent with American Fisheries Society guidelines (<https://fisheries.org/docs/wp/Guidelines-for-Use-of-Fishes.pdf>). The sampled species do not have protected status and the permit does not require a specific killing procedure. Animals were never held in captivity and no experiments were done with live animals. Takings for scientific research are equivalent to commercial harvest or recreational angling and do not fall under IACUC review for euthanasia protocol (Yanong RP, Hartman KH, Watson CA, Hill JE, Petty BD, Francis-Floyd R. Fish slaughter, killing, and euthanasia: a review of major published US guidance documents and general considerations of methods. EDIS. 2007 Aug 30;2007(18)).

### Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

All activities were performed by state and federal employees authorized to do so by their agencies. Sampling was authorized under a State of Missouri Wildlife Collector's Permit (# 19224) issued to R. Hrabik.

### Data Availability

The following information was supplied regarding data availability:

The raw short-read data for this study is available at NCBI: [PRJNA806159](#) and [PRJNA806850](#).

The newly assembled mitogenomes for the *P. tenellus* and *P. notatus* samples are available at NCBI: [OM718772–OM718774](#) and [OM743017](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13954#supplemental-information>.

## REFERENCES

- Anderson JL, Rodríguez Mari A, Braasch I, Amores A, Hohenlohe P, Batzel P, Postlethwait JH. 2012. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLOS ONE* 7:e40701 DOI 10.1371/journal.pone.0040701.
- Angata T. 2006. Molecular diversity and evolution of the Siglec family of cell–surface lectins. *Molecular diversity* 10:555–566 DOI 10.1007/s11030-006-9029-1.
- Ankley GT, Villeneuve DL. 2006. The fathead minnow in aquatic toxicology: past, present and future. *Aquatic Toxicology* 78:91–102 DOI 10.1016/j.aquatox.2006.01.018.
- Baird DJ, Barata C. 1998. Genetic variation in the response of *Daphnia* to toxic substances: implications. In: Forbes VE, ed. *Genetics and ecotoxicology*. Philadelphia: Taylor & Francis, 207–222.
- Ballesteros-Nova NE, Pérez-Rodríguez R, Beltrán-López RG, Domínguez-Domínguez O. 2019. Genetic differentiation in the southern population of the fathead minnow *Pimephales promelas* Rafinesque (Actinopterygii: Cyprinidae). *PeerJ* 7:e6224 DOI 10.7717/peerj.6224.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19:455–477 DOI 10.1089/cmb.2012.0021.
- Beggel S, Werner I, Cannon RE, Geist JP. 2012. Impacts of the phenylpyrazole insecticide fipronil on larval fish: time-series gene transcription responses in fathead minnow (*Pimephales promelas*) following short-term exposure. *Science of the Total Environment* 426:160–165 DOI 10.1016/j.scitotenv.2012.04.005.

- Benoit D, Puglisi F, Olson D. 1982.** A fathead minnow *Pimephales promelas* early life stage toxicity test method evaluation and exposure to four organic chemicals. *Environmental Pollution Series A, Ecological and Biological* **28**:189–197 DOI [10.1016/0143-1471\(82\)90075-6](https://doi.org/10.1016/0143-1471(82)90075-6).
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsich G, Pütz J, Middelndorf M, Stadler PF. 2013.** MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**:313–319 DOI [10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023).
- Bielawski JP, Brault A, Gold J. 2002.** Phylogenetic relationships within the genus *Pimephales* as inferred from ND4 and ND4L nucleotide sequences. *Journal of Fish Biology* **61**:293–297 DOI [10.1111/j.1095-8649.2002.tb01754.x](https://doi.org/10.1111/j.1095-8649.2002.tb01754.x).
- Blanton RE, Page LM, Ennis BM. 2011.** Phylogenetic relationships of *Opsopoeodus emiliae*, with comments on the taxonomic implications of discordance among datasets. *Copeia* **2011**:82–92 DOI [10.1643/CI-10-004](https://doi.org/10.1643/CI-10-004).
- Boitard S, Kofler R, Françoise P, Robelin D, Schlötterer C, Futschik A. 2013.** Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Molecular Ecology Resources* **13**:337–340 DOI [10.1111/1755-0998.12063](https://doi.org/10.1111/1755-0998.12063).
- Brown AR, Bickley LK, Ryan TA, Paull GC, Hamilton PB, Owen SF, Sharpe AD, Tyler CR. 2012.** Differences in sexual development in inbred and outbred zebrafish (*Danio rerio*) and implications for chemical testing. *Aquatic Toxicology* **112**:27–38 DOI [10.1016/j.aquatox.2012.01.017](https://doi.org/10.1016/j.aquatox.2012.01.017).
- Burns FR, Cogburn AL, Ankley GT, Villeneuve DL, Waits E, Chang Y-J, Llaca V, Deschamps SD, Jackson RE, Hoke RA. 2016.** Sequencing and de novo draft assemblies of a fathead minnow (*Pimephales promelas*) reference genome. *Environmental Toxicology and Chemistry* **35**:212–217 DOI [10.1002/etc.3186](https://doi.org/10.1002/etc.3186).
- Calkins JD, El-Hinn D, Swanson WJ. 2007.** Adaptive evolution in an avian reproductive protein: ZP3. *Journal of Molecular Evolution* **65**:555–563 DOI [10.1007/s00239-007-9034-8](https://doi.org/10.1007/s00239-007-9034-8).
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* **10**:1–9 DOI [10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421).
- Campos HH, Hubbs C. 1973.** Taxonomic implications of the karyotype of *Opsopoeodus emiliae*. *Copeia* **1973**:161–163 Available at <https://www.jstor.org/stable/1442381> DOI [10.2307/1442381](https://doi.org/10.2307/1442381).
- Cannons JL, Tangye SG, Schwartzberg PL. 2011.** SLAM family receptors and SAP adaptors in immunity. *Annual Review of Immunology* **29**:665–705 DOI [10.1146/annurev-immunol-030409-101302](https://doi.org/10.1146/annurev-immunol-030409-101302).
- Catchen JM, Braasch I, Postlethwait JH. 2011.** Conserved synteny and the zebrafish genome. *Methods in Cell Biology* **104**:259–285 DOI [10.1016/B978-0-12-374814-0.00015-X](https://doi.org/10.1016/B978-0-12-374814-0.00015-X).
- Chen Z, Omori Y, Koren S, Shirokiya T, Kuroda T, Miyamoto A, Wada H, Fujiyama A, Toyoda A, Zhang S, Wolfsberg TG, Kawakami K, Phillippy AM, Mullikin JC,**

- Burgess SM. 2019.** De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. *Science Advances* 5:eaav0547 DOI 10.1126/sciadv.aav0547.
- Cheverud JM, Rutledge J, Atchley WR. 1983.** Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution* 37:895–905 DOI 10.2307/2408405.
- Chuang T, Xu X, Kaartinen V, Heisterkamp N, Groffen J, Bokoch G. 1995.** Abr and Bcr are multifunctional regulators of the Rho GTP-binding protein family. *Proceedings of the National Academy of Sciences of the United States of America* 92:10282–10286 DOI 10.1073/pnas.92.22.10282.
- Clark EA, Giltiay NV. 2018.** CD22: a regulator of innate and adaptive B cell responses and autoimmunity. *Frontiers in Immunology* 9:2235 DOI 10.3389/fimmu.2018.02235.
- Cornman RS, Klymus KE, Thompson NL, Hrabik R. 2022.** Genomic variation in the genus *Pimephales*: raw sequence data and single-nucleotide polymorphisms. US Geological Survey data release, XEUNR DOI 10.5066/P9X.
- Corral-Serrano JC, Lamers IJC, van Reeuwijk J, Duijkers L, Hoogendoorn ADM, Yildirim A, Argyrou N, Ruigrok RAA, Letteboer SJF, Butcher R, Van Essen MD, Sakami S, Van Beersum SEC, Palczewski K, Cheetham ME, Liu Q, Boldt K, Wolfrum U, Ueffing M, Garanto A, Roepman R, Collin RWJ. 2020.** PCARE and WASF3 regulate ciliary F-actin assembly that is required for the initiation of photoreceptor outer segment disk formation. *Proceedings of the National Academy of Sciences of the United States of America* 117:9922–9931 DOI 10.1073/pnas.1903125117.
- Crawford DL, Schulte PM, Whitehead A, Oleksiak M. 2020.** Evolutionary physiology and genomics in the highly adaptable killifish (*Fundulus heteroclitus*). *Comprehensive Physiology* 10:637–671 DOI 10.1002/cphy.c190004.
- Criscitiello MF, Wermenstam NE, Pilström L, McKinney EC. 2004.** Allelic polymorphism of T-cell receptor constant domains is widespread in fishes. *Immunogenetics* 55:818–824 DOI 10.1007/s00251-004-0652-7.
- Cureton JC, Broughton RE. 2014.** Rapid morphological divergence of a stream fish in response to changes in water flow. *Biology Letters* 10:20140352 DOI 10.1098/rsbl.2014.0352.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, De Pristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011.** The variant call format and VCFtools. *Bioinformatics* 27:2156–2158 DOI 10.1093/bioinformatics/btr330.
- Denny JS. 1987.** Guidelines for the culture of fathead minnows (*Pimephales promelas*) for use in toxicity tests. US Environmental Protection Agency EPA/600/3-87/001.
- Diekmann H, Stuermer CA. 2009.** Zebrafish neurolin-a and-b, orthologs of ALCAM, are involved in retinal ganglion cell differentiation and retinal axon pathfinding. *Journal of Comparative Neurology* 513:38–50 DOI 10.1002/cne.21928.
- Dugal L, Thomas L, Jensen MR, Sigsgaard EE, Simpson T, Jarman S, Thomsen PF, Meekan M. 2022.** Individual haplotyping of whale sharks from seawater environmental DNA. *Molecular Ecology Resources* 22:56–65 DOI 10.1111/1755-0998.13451.

- Ejarque-Ortiz A, Solà C, Martínez-Barriocanal Á, Schwartz Jr S, Martín M, Peluffo H, Sayós J. 2015.** The receptor CMRF35-like molecule-1 (CLM-1) enhances the production of LPS-induced pro-inflammatory mediators during microglial activation. *PLOS ONE* **10**:e0123928 DOI [10.1371/journal.pone.0123928](https://doi.org/10.1371/journal.pone.0123928).
- Fang B, Kempainen P, Momigliano P, Feng X, Merilä J. 2020.** On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution* **4**:1105–1115 DOI [10.1038/s41559-020-1222-6](https://doi.org/10.1038/s41559-020-1222-6).
- Ferrero-Serrano Á, Assmann SM. 2019.** Phenotypic and genome-wide association with the local environment of *Arabidopsis*. *Nature Ecology & Evolution* **3**:274–285 DOI [10.1038/s41559-018-0754-5](https://doi.org/10.1038/s41559-018-0754-5).
- Ferretti L, Ramos-Onsins SE, Pérez-Enciso M. 2013.** Population genomics from pool sequencing. *Molecular Ecology* **22**:5561–5576 DOI [10.1111/mec.12522](https://doi.org/10.1111/mec.12522).
- Feulner PG, Chain FJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, Mundry M, Samonte IE, Stoll M, Milinski M, Reusch TBH, Bornberg-Bauer E. 2013.** Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology* **22**:635–649 DOI [10.1111/j.1365-294X.2012.05680.x](https://doi.org/10.1111/j.1365-294X.2012.05680.x).
- Fijarczyk A, Babik W. 2015.** Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology* **24**:3529–3545 DOI [10.1111/mec.13226](https://doi.org/10.1111/mec.13226).
- Fioletov VE, Kimlin MG, Krotkov N, McArthur LB, Kerr JB, Wardle DI, Herman JR, Meltzer R, Mathews TW, Kaurola J. 2004.** UV index climatology over the United States and Canada from ground-based and satellite estimates. *Journal of Geophysical Research: Atmospheres* **109**: DOI [10.1029/2004JD004820](https://doi.org/10.1029/2004JD004820).
- Flanagan SP, Jones AG. 2017.** Constraints on the FST–heterozygosity outlier approach. *Journal of Heredity* **108**:561–573 DOI [10.1093/jhered/esx048](https://doi.org/10.1093/jhered/esx048).
- Freyhof J, Lieckfeldt D, Pitra C, Ludwig A. 2005.** Molecules and morphology: evidence for introgression of mitochondrial DNA in Dalmatian cyprinids. *Molecular Phylogenetics and Evolution* **37**:347–354 DOI [10.1016/j.ympev.2005.07.018](https://doi.org/10.1016/j.ympev.2005.07.018).
- Fritz G, Kaina B. 1997.** rhoB encoding a UV-inducible Ras-related small GTP-binding protein is regulated by GTPases of the Rho family and independent of JNK, ERK, and p38 MAP kinase. *Journal of Biological Chemistry* **272**:30637–30644 DOI [10.1074/jbc.272.49.30637](https://doi.org/10.1074/jbc.272.49.30637).
- Fritz G, Kaina B, Aktories K. 1995.** The Ras-related small GTP-binding protein RhoB is immediate-early inducible by DNA damaging treatments. *Journal of Biological Chemistry* **270**:25172–25177 DOI [10.1074/jbc.270.42.25172](https://doi.org/10.1074/jbc.270.42.25172).
- Futschik A, Schlötterer C. 2010.** The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* **186**:207–218 DOI [10.1534/genetics.110.114397](https://doi.org/10.1534/genetics.110.114397).
- Gautier M. 2015.** Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**:1555–1579 DOI [10.1534/genetics.115.181453](https://doi.org/10.1534/genetics.115.181453).
- Goodman CA. 2013.** The role of mTORC1 in regulating protein synthesis and skeletal muscle mass in response to various mechanical stimuli. *Reviews of Physiology, Biochemistry and Pharmacology* **166**:43–95 DOI [10.1007/112\\_2013\\_17](https://doi.org/10.1007/112_2013_17).

- Haas TC, Heins DC, Blum MJ. 2015. Predictors of body shape among populations of a stream fish (*Cyprinella venusta*, Cypriniformes: Cyprinidae). *Biological Journal of the Linnean Society* 115:842–858 DOI 10.1111/bij.12539.
- Hamann J, Hsiao C-C, Lee CS, Ravichandran KS, Lin H-H. 2016. Adhesion GPCRs as modulators of immune cell function. In: Langenhan T, Schöneberg T, eds. *Adhesion G protein-coupled receptors. Handbook of experimental pharmacology*. vol 234. Cham: Springer, 329–350 DOI 10.1007/978-3-319-41523-9\_15.
- Hay IM, Fearnley GW, Rios P, Köhn M, Sharpe HJ, Deane JE. 2020. The receptor PTPRU is a redox sensitive pseudophosphatase. *Nature Communications* 11:1–13 DOI 10.1038/s41467-020-17076-w.
- Hoffmann AA, Hallas R, Sinclair C, Partridge L. 2001. Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. *Evolution* 55:436–438 DOI 10.1111/j.0014-3820.2001.tb01305.x.
- Holman LE, Hollenbeck CM, Ashton TJ, Johnston IA. 2019. Demonstration of the use of environmental DNA for the non-invasive genotyping of a bivalve mollusk, the European flat oyster (*Ostrea edulis*). *Frontiers in Genetics* 10:1159 DOI 10.3389/fgene.2019.01159.
- Horzmann KA, Freeman JL. 2018. Making waves: new developments in toxicology with the zebrafish. *Toxicological Sciences* 163:5–12 DOI 10.1093/toxsci/kfy044.
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel J, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Lovel J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliott D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthavadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper JD, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SMJ, Enright A, Geisler R, Plasterk RHA, Lee C, Westerfield M, Jong PJd, Zon LI, Postlethwait JH, Nüsslein-Volhard C,

- Hubbard TJP, Crollius HR, Rogers J, Stemple DL. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**:498–503 DOI [10.1038/nature12111](https://doi.org/10.1038/nature12111).
- Hu YW, Yu ZL, Xue NN, Nie P, Chang MX. 2014. Expression and protective role of two novel NACHT-containing proteins in pathogen infection. *Developmental & Comparative Immunology* **46**:323–332 DOI [10.1016/j.dci.2014.05.007](https://doi.org/10.1016/j.dci.2014.05.007).
- Huang Y, Feulner PG, Eizaguirre C, Lenz TL, Bornberg-Bauer E, Milinski M, Reusch TB, Chain FJ. 2019. Genome-wide genotype-expression relationships reveal both copy number and single nucleotide differentiation contribute to differential gene expression between stickleback ecotypes. *Genome Biology and Evolution* **11**:2344–2359 DOI [10.1093/gbe/evz148](https://doi.org/10.1093/gbe/evz148).
- Hubbs CL, Black JD. 1947. Revision of *Ceratichthys*, a genus of American cyprinid fishes. In: *Museum of Zoology publication number 66*. Ann Arbor: University of Michigan Press.
- Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167–170 DOI [10.1038/335167a0](https://doi.org/10.1038/335167a0).
- Hughes AL, Nei M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America* **86**:958–962 DOI [10.1073/pnas.86.3.958](https://doi.org/10.1073/pnas.86.3.958).
- Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M, Nishida M. 2013. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution* **30**:2531–2540 DOI [10.1093/molbev/mst141](https://doi.org/10.1093/molbev/mst141).
- Jacquemin SJ, Pyron M. 2016. A century of morphological variation in Cyprinidae fishes. *BMC Ecology* **16**:1–18 DOI [10.1186/s12898-016-0104-x](https://doi.org/10.1186/s12898-016-0104-x).
- Jansen JG, Tsaalbi-Shtylik A, Hendriks G, Verspuy J, Gali H, Haracska L, de Wind N. 2009. Mammalian polymerase  $\zeta$  is essential for post-replication repair of UV-induced DNA lesions. *DNA Repair* **8**:1444–1451 DOI [10.1016/j.dnarep.2009.09.006](https://doi.org/10.1016/j.dnarep.2009.09.006).
- Jansen R. 2004. Quantitative trait loci in inbred lines. In: *Handbook of statistical genetics*. Wiley Online Library DOI [10.1002/0470022620.bbc16](https://doi.org/10.1002/0470022620.bbc16).
- Jongbloets BC, Pasterkamp RJ. 2014. Semaphorin signalling during development. *Development* **141**:3292–3297 DOI [10.1242/dev.105544](https://doi.org/10.1242/dev.105544).
- Kansakar U, Wang W, Markovic V, Sossey-Alaoui K. 2020. Elucidating the molecular signaling pathways of WAVE3. *Annals of Translational Medicine* **8**:900 DOI [10.21037/atm.2020.02.16](https://doi.org/10.21037/atm.2020.02.16).
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**:3059–3066 DOI [10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436).
- Kern EM, Langerhans RB. 2018. Urbanization drives contemporary evolution in stream fish. *Global Change Biology* **24**:3791–3803 DOI [10.1111/gcb.14115](https://doi.org/10.1111/gcb.14115).

- Kimura S. 2018.** Molecular insights into the mechanisms of M-cell differentiation and transcytosis in the mucosa-associated lymphoid tissues. *Anatomical Science International* **93**:23–34 DOI [10.1007/s12565-017-0418-6](https://doi.org/10.1007/s12565-017-0418-6).
- Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011.** PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLOS ONE* **6**:e15925 DOI [10.1371/journal.pone.0015925](https://doi.org/10.1371/journal.pone.0015925).
- Kofler R, Pandey RV, Schlötterer C. 2011.** PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**:3435–3436 DOI [10.1093/bioinformatics/btr589](https://doi.org/10.1093/bioinformatics/btr589).
- Kosakovsky-Pond SL, Muse SV. 2005.** HyPhy: hypothesis testing using phylogenies. In: Nielsen R, ed. *Statistical methods in molecular evolution*. New York: Springer, 125–181.
- Krogh A, Larsson B, Heijne GVon, Sonnhammer EL. 2001.** Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**:567–580 DOI [10.1006/jmbi.2000.4315](https://doi.org/10.1006/jmbi.2000.4315).
- Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M, Snyder M. 2014.** Whole-genome haplotyping using long reads and statistical methods. *Nature Biotechnology* **32**:261–266 DOI [10.1038/nbt.2833](https://doi.org/10.1038/nbt.2833).
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018.** MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**:1547–1549 DOI [10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096).
- Kusmic C, Gualtieri P. 2000.** Morphology and spectral sensitivities of retinal and extraretinal photoreceptors in freshwater teleosts. *Micron* **31**:183–200 DOI [10.1016/S0968-4328\(99\)00081-5](https://doi.org/10.1016/S0968-4328(99)00081-5).
- Laing KJ, Purcell MK, Winton JR, Hansen JD. 2008.** A genomic view of the NOD-like receptor family in teleost fish: identification of a novel NLR subfamily in zebrafish. *BMC Evolutionary Biology* **8**:1–15 DOI [10.1186/1471-2148-8-42](https://doi.org/10.1186/1471-2148-8-42).
- Lange SS, Tomida J, Boulware KS, Bhetawal S, Wood RD. 2016.** The polymerase activity of mammalian DNA pol  $\zeta$  is specifically required for cell and embryonic viability. *PLOS Genetics* **12**:e1005759 DOI [10.1371/journal.pgen.1005759](https://doi.org/10.1371/journal.pgen.1005759).
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359 DOI [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).
- Larrick SR, Dickson K, Cherry D, Cairns J. 1978.** Determining fish avoidance of polluted water. *Hydrobiologia* **61**:257–265 DOI [10.1007/BF00044451](https://doi.org/10.1007/BF00044451).
- Latta RG. 2004.** Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* **161**:51–58 DOI [10.1046/j.1469-8137.2003.00920.x](https://doi.org/10.1046/j.1469-8137.2003.00920.x).
- Latter BD, Mulley JC. 1995.** Genetic adaptation to captivity and inbreeding depression in small laboratory populations of *Drosophila melanogaster*. *Genetics* **139**:255–66 DOI [10.1093/genetics/139.1.255](https://doi.org/10.1093/genetics/139.1.255).

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079 DOI [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).
- Lin Y-C, Owen N, Minko IG, Lange SS, Tomida J, Li L, Stone MP, Wood RD, McCullough AK, Lloyd RS. 2016. DNA polymerase  $\zeta$  limits chromosomal damage and promotes cell survival following aflatoxin exposure. *Proceedings of the National Academy of Sciences of the United States of America* 113:13774–13779 DOI [10.1073/pnas.1609024113](https://doi.org/10.1073/pnas.1609024113).
- Liu GY, Sabatini DM. 2020. mTOR at the nexus of nutrition, growth, ageing and disease. *Nature Reviews Molecular Cell Biology* 21:183–203 DOI [10.1038/s41580-019-0199-y](https://doi.org/10.1038/s41580-019-0199-y).
- Macqueen DJ, Kristjansson BK, Paxton CG, Vieira VL, Johnston IA. 2011. The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. *Molecular Ecology* 20:3167–3184 DOI [10.1111/j.1365-294X.2011.05172.x](https://doi.org/10.1111/j.1365-294X.2011.05172.x).
- Mamouni K, Cristini A, Guirouilh-Barbat J, Monferran S, Lemarié A, Faye J-C, Lopez BS, Favre G, Sordet O. 2014. RhoB promotes  $\gamma$ H2AX dephosphorylation and DNA double-strand break repair. *Molecular and Cellular Biology* 34:3144–3155 DOI [10.1128/MCB.01525-13](https://doi.org/10.1128/MCB.01525-13).
- Manor ML, Cleveland BM, Kenney PBrett, Yao J, Leeds T. 2015. Differences in growth, fillet quality, and fatty acid metabolism-related gene expression between juvenile male and female rainbow trout. *Fish Physiology and Biochemistry* 41:533–547 DOI [10.1007/s10695-015-0027-z](https://doi.org/10.1007/s10695-015-0027-z).
- Martinson JW, Bencic DC, Toth GP, Kostich MS, Flick RW, See MJ, Lattier D, Biales AD, Huang W. 2021. De novo assembly of the nearly complete fathead minnow reference genome reveals a repetitive but compact genome. *Environmental Toxicology and Chemistry* 41:448–461 DOI [10.1002/etc.5266](https://doi.org/10.1002/etc.5266).
- Milano I, Babbucci M, Cariani A, Atanassova M, Bekkevold D, Carvalho GR, Espineira M, Fiorentino F, Garofalo G, Geffen AJ, Hansen JH, Helyar SJ, Nielsen EE, Ogden R, Patarnello T, Stagioni M, ISHPOPTRACE Consortium, Tinti F, Bargelloni L. 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology* 23:118–135 DOI [10.1111/mec.12568](https://doi.org/10.1111/mec.12568).
- Moshkovits I, Shik D, Itan M, Karo-Atar D, Bernshtein B, Hershko A, van Lookeren Campagne M, Munitz A. 2014. CMRF35-like molecule 1 (CLM-1) regulates eosinophil homeostasis by suppressing cellular chemotaxis. *Mucosal Immunology* 7:292–303 DOI [10.1038/mi.2013.47](https://doi.org/10.1038/mi.2013.47).
- Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, Scheffler K, Kosakovsky Pond SL. 2015. Gene-wide identification of episodic selection. *Molecular Biology and Evolution* 32:1365–1371 DOI [10.1093/molbev/msv035](https://doi.org/10.1093/molbev/msv035).
- Nico L, Fuller P, Neilson M. 2017. Pimephales promelas. Available at <https://nas.er.usgs.gov/queries/factsheet.aspx?SpeciesID=621>.

- Palmai-Pallag T, Khodabukus N, Kinarsky L, Leir SH, Sherman S, Hollingsworth MA, Harris A. 2005.** The role of the SEA (sea urchin sperm protein, enterokinase and agrin) module in cleavage of membrane-tethered mucins. *The FEBS Journal* 272:2901–2911 DOI [10.1111/j.1742-4658.2005.04711.x](https://doi.org/10.1111/j.1742-4658.2005.04711.x).
- Perteua G, Perteua M. 2020.** GFF utilities: GffRead and GffCompare. *F1000Research* 9:304 DOI [10.12688/f1000research.23297.2](https://doi.org/10.12688/f1000research.23297.2).
- Pickering QH. 1988.** Evaluation and comparison of two short-term fathead minnow tests for estimating chronic toxicity. *Water Research* 22:883–893 DOI [10.1016/0043-1354\(88\)90026-7](https://doi.org/10.1016/0043-1354(88)90026-7).
- Pollard DA, Iyer VN, Moses AM, Eisen MB. 2006.** Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLOS Genetics* 2:e173 DOI [10.1371/journal.pgen.0020173](https://doi.org/10.1371/journal.pgen.0020173).
- Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. 2008.** Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* 179:555–567 DOI [10.1534/genetics.107.083006](https://doi.org/10.1534/genetics.107.083006).
- R Core Team. 2018.** R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Reid NM, Jackson CE, Gilbert D, Minx P, Montague MJ, Hampton TH, Helfrich LW, King BL, Nacci DE, Aluru N, Karchner SI, Colbourne JK, Hahn ME, Shaw JR, Oleksiak MF, Crawford DL, Warren WC, Whitehead A. 2017.** The landscape of extreme genomic variation in the highly adaptable Atlantic killifish. *Genome Biology and Evolution* 9:659–676 DOI [10.1093/gbe/evx023](https://doi.org/10.1093/gbe/evx023).
- Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, Shaw JR, Karchner SI, Hahn ME, Nacci D, Oleksiak MF, Crawford DL. 2016.** The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* 354:1305–1308 DOI [10.1126/science.aah4993](https://doi.org/10.1126/science.aah4993).
- Ronquist F, Huelsenbeck JP. 2003.** MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574 DOI [10.1093/bioinformatics/btg180](https://doi.org/10.1093/bioinformatics/btg180).
- Ross ST. 2001.** *Inland fishes of Mississippi*. University Press of Mississippi. Available at [https://aquila.usm.edu/faculty\\_books/16](https://aquila.usm.edu/faculty_books/16) (accessed on 18 July 2022).
- Sahoo BR. 2020.** Structure of fish Toll-like receptors (TLR) and NOD-like receptors (NLR). *International Journal of Biological Macromolecules* 161:1602–1617 DOI [10.1016/j.ijbiomac.2020.07.293](https://doi.org/10.1016/j.ijbiomac.2020.07.293).
- Schmidt TR, Dowling TE, Gold JR. 1994.** Molecular systematics of the genus *Pimephales* (Teleostei: Cyprinidae). *The Southwestern Naturalist* 241–248 DOI [10.2307/3671588](https://doi.org/10.2307/3671588).
- Schönhuth S, Beachum CE, Knouft JH, Mayden RL. 2016.** Phylogeny and genetic variation within the widely distributed bluntnose minnow, *Pimephales notatus* (Cyprinidae), in North America. *Zootaxa* 4168:38–60 DOI [10.11646/zootaxa.4168.1.2](https://doi.org/10.11646/zootaxa.4168.1.2).
- Schroeter JC, Maloy AP, Rees CB, Bartron ML. 2020.** Fish mitochondrial genome sequencing: expanding genetic resources to support species detection and biodiversity

- monitoring using environmental DNA. *Conservation Genetics Resources* 12:433–446 DOI 10.1007/s12686-019-01111-0.
- Seiliez I, Gabillard J-C, Skiba-Cassy S, Garcia-Serrana D, Gutiérrez J, Kaushik S, Panserat S, Tesseraud S. 2008.** An in vivo and in vitro assessment of TOR signaling cascade in rainbow trout (*Oncorhynchus mykiss*). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 295:R329–R335 DOI 10.1152/ajpregu.00146.2008.
- Shik D, Moshkovits I, Karo-Atar D, Reichman H, Munitz A. 2014.** Interleukin-33 requires CMRF 35-like molecule-1 expression for induction of myeloid cell activation. *Allergy* 69:719–729 DOI 10.1111/all.12388.
- Shum BP, Guethlein L, Flodin LR, Adkison MA, Hedrick RP, Nehring RB, Stet RJ, Secombes C, Parham P. 2001.** Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *The Journal of Immunology* 166:3297–3308 DOI 10.4049/jimmunol.166.5.3297.
- Skiba-Cassy S, Lansard M, Panserat S, Médale F. 2009.** Rainbow trout genetically selected for greater muscle fat content display increased activation of liver TOR signaling and lipogenic gene expression. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 297:R1421–R1429 DOI 10.1152/ajpregu.00312.2009.
- Slade R, McCallum H. 1992.** Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* 132:861 DOI 10.1093/genetics/132.3.861.
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015.** Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution* 32:1342–1353 DOI 10.1093/molbev/msv022.
- Steffen LS, Guyon JR, Vogel ED, Howell MH, Zhou Y, Weber GJ, Zon LI, Kunkel LM. 2007.** The zebrafish runzel muscular dystrophy is linked to the titin gene. *Developmental Biology* 309:180–192 DOI 10.1016/j.ydbio.2007.06.015.
- Stinson SA, Hasenbein S, Connon RE, Deng X, Alejo JS, Lawler SP, Holland EB. 2022.** Agricultural surface water, imidacloprid, and chlorantraniliprole result in altered gene expression and receptor activation in *Pimephales promelas*. *Science of The Total Environment* 806:150920 DOI 10.1016/j.scitotenv.2021.150920.
- Storz JF. 2005.** Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14:671–688 DOI 10.1111/j.1365-294X.2005.02437.x.
- Su S, Li H, Du F, Zhang C, Li X, Jing X, Liu L, Li Z, Yang X, Xu P, Yuan X, Zhu J, Bouzoualegh R. 2018.** Combined QTL and genome scan analyses with the help of 2b-RAD identify growth-associated genetic markers in a new fast-growing carp strain. *Frontiers in Genetics* 9:592 DOI 10.3389/fgene.2018.00592.
- Tajima F. 1983.** Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460 DOI 10.1093/genetics/105.2.437.
- Tong C, Li M. 2020.** Genomic signature of accelerated evolution in a saline-alkaline lake-dwelling Schizothoracine fish. *International Journal of Biological Macromolecules* 149:341–347 DOI 10.1016/j.ijbiomac.2020.01.207.

- Turner LM, Hoekstra HE. 2008.** Reproductive protein evolution within and between species: maintenance of divergent ZP3 alleles in *Peromyscus*. *Molecular Ecology* 17:2616–2628 DOI [10.1111/j.1365-294X.2008.03780.x](https://doi.org/10.1111/j.1365-294X.2008.03780.x).
- Uchii K, Doi H, Minamoto T. 2016.** A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. *Molecular Ecology Resources* 16:415–422 DOI [10.1111/1755-0998.12460](https://doi.org/10.1111/1755-0998.12460).
- Ugwu SI, Shiba K, Inaba K, Morita M. 2018.** A unique seminal plasma protein, zona pellucida 3-like protein, has Ca<sup>2+</sup>-dependent sperm agglutination activity. *Zoological Science* 35:161–171 DOI [10.2108/zs170150](https://doi.org/10.2108/zs170150).
- Valente LM, Moutou KA, Conceicao LE, Engrola S, Fernandes JM, Johnston IA. 2013.** What determines growth potential and juvenile quality of farmed fish species? *Reviews in Aquaculture* 5:S168–S193 DOI [10.1111/raq.12020](https://doi.org/10.1111/raq.12020).
- Vega FM, Ridley AJ. 2018.** The RhoB small GTPase in physiology and disease. *Small GTPases* 9:384–393 DOI [10.1080/21541248.2016.1253528](https://doi.org/10.1080/21541248.2016.1253528).
- Wang K, McCarter R, Wright J, Beverly J, Ramirez-Mitchell R. 1991.** Regulation of skeletal muscle stiffness and elasticity by titin isoforms: a test of the segmental extension model of resting tension. *Proceedings of the National Academy of Sciences of the United States of America* 88:7101–7105 DOI [10.1073/pnas.88.16.7101](https://doi.org/10.1073/pnas.88.16.7101).
- Weetman D, Wilding CS, Neafsey DE, Müller P, Ochomo E, Isaacs AT, Steen K, Rippon EJ, Morgan JC, Mawejje HD, Rigden DJ. 2018.** Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African *Anopheles gambiae*. *Scientific Reports* 8:1–2 DOI [10.1038/s41598-018-21265-5](https://doi.org/10.1038/s41598-018-21265-5).
- Wei L, Liu Y, Dubchak I, Shon J, Park J. 2002.** Comparative genomics approaches to study organism similarities and differences. *Journal of Biomedical Informatics* 35:142–150 DOI [10.1016/S1532-0464\(02\)00506-3](https://doi.org/10.1016/S1532-0464(02)00506-3).
- Whitehead A, Clark BW, Reid NM, Hahn ME, Nacci D. 2017.** When evolution is the solution to pollution: key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evolutionary Applications* 10:762–783 DOI [10.1111/eva.12470](https://doi.org/10.1111/eva.12470).
- Whitehead A, Triant DA, Champlin D, Nacci D. 2010.** Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology* 19:5186–203 DOI [10.1111/j.1365-294X.2010.04829.x](https://doi.org/10.1111/j.1365-294X.2010.04829.x).
- Whittaker CA, Hynes RO. 2002.** Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Molecular Biology of the Cell* 13:3369–3387 DOI [10.1091/mbc.e02-05-0259](https://doi.org/10.1091/mbc.e02-05-0259).
- Wilson CA, High SK, McCluskey BM, Amores A, Yan Y, Titus TA, Anderson JL, Batzel P, CarvanIII MJ, Schartl M, Postlethwait JH. 2014.** Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* 198:1291–1308 DOI [10.1534/genetics.114.169284](https://doi.org/10.1534/genetics.114.169284).
- Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, Zhang Y, Huo L, Zhao Z, Cao D, Lu C, Li C, Zhou Y, Liu Z, Fan Z, Shan G, Li X, Wu S, Song L, Hou G, Jiang Y, Jeney Z, Yu D, Wang L, Shao C, Song L, Sun J, Ji P, Wang**

- J, Li Q, Xu L, Sun F, Feng J, Wang C, Wang S, Wang B, Li Y, Zhu Y, Xue W, Zhao L, Wang J, Gu Y, Lv W, Wu K, Xiao J, Wu J, Zhang Z, Yu J, Sun X. 2014. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics* 46:1212–1219 DOI 10.1038/ng.3098.
- Yang J, Chen X, Bai J, Fang D, Qiu Y, Jiang W, Yuan H, Bian C, Lu J, He S, Pan X, Zhang Y, Wang X, You X, Wang Y, Sun Y, Mao D, Liu Y, Fan G, Zhang H, Chen X, Zhang X, Zheng L, Wang J, Cheng L, Chen J, Ruan Z, Li J, Yu H, Peng C, Ma X, Xu J, He Y, Xu Z, Xu P, Wang J, Yang H, Wang J, Whitten T, Xu X, Shi Q. 2016. The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation. *BMC Biology* 14:1–13 DOI 10.1186/s12915-015-0223-4.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24:1586–1591 DOI 10.1093/molbev/msm088.
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* 17:32–43 DOI 10.1093/oxfordjournals.molbev.a026236.
- You J-S, McNally RM, Jacobs BL, Privett RE, Gundermann DM, Lin K-H, Steinert ND, Goodman CA, Hornberger TA. 2019. The role of raptor in the mechanical load-induced regulation of mTOR signaling, protein synthesis, and skeletal muscle hypertrophy. *The FASEB Journal* 33:4021–4034 DOI 10.1096/fj.201801653RR.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* 11:1–12 DOI 10.1186/gb-2010-11-2-r14.
- Yuen M, Ottenheijm CA. 2020. Nebulin: big protein with big responsibilities. *Journal of Muscle Research and Cell Motility* 41:103–124 DOI 10.1007/s10974-019-09565-3.
- Zhang S, Cui P. 2014. Complement system in zebrafish. *Developmental & Comparative Immunology* 46:3–10 DOI 10.1016/j.dci.2014.01.010.