CaveCrawler: An interactive analysis suite for cavefish bioinformatics

Annabel Perry[1], Suzanne E. McGaugh[2], Alex C. Keene[1#], and Heath Blackmon[1#]

1. Department of Biology, Texas A&M University, College Station, TX
2. Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, Minnesota, United States of America.

address correspondence to hblackmon@bio.tamu.edu and akeene@bio.tamu.edu

Abstract

The growing use of genomics data in diverse animal models provides the basis for identifying genomic and transcriptional differences across species and contexts. Databases containing genomic and functional data have played critical roles in the development of numerous genetic models but are lacking for most emerging models of evolution. There is a rapidly expanding use of genomic, transcriptional, and functional genetic approaches to study diverse traits of the Mexican tetra, *Astyanax mexicanus*. This species exists as two morphs, eyed surface populations and at least 30 blind cave populations, providing a system to study convergent evolution. We have generated a web-based analysis suite that integrates datasets from different studies to identify how gene transcription and genetic markers of selection differ between populations and across experimental contexts. Results can be processed with other analysis platforms including Gene Ontology (GO) to enable biological inference from cross-study patterns and identify future avenues of research. Furthermore, the framework that we have built *A. mexicanus* can readily applied to other emerging model systems.

Introduction

The reduced cost and increased efficiency of sequencing has led to enormous growth in the application of sequencing approaches to study diverse biological processes. In previous decades, these approaches were predominantly performed on a small number of genetically amendable model organisms including *Caenorhabditis elegans*, *Drosophila melanogaster*, zebrafish, and mouse. Model-organism-specific databases have been generated for each of these model systems, providing critical resources that decrease access barriers to genomic and phenotypic data (1-3). Recently, there has been increased application of genomic and molecular approaches to non-standard model systems, as these model systems may enable comparative evolutionary studies not possible in traditional systems (4). However, a lack of databases and analytic tools for many of these emerging model organisms impedes analysis of genomic data collected across different studies.

The Mexican tetra, *Astyanax mexicanus* is an emerging model system to study the convergent evolution of diverse biological traits. These fish are comprised of a single population of river dwelling surface fish and at least 30 cavefish populations of the same species (5). *Astyanax mexicanus* cavefish populations have independently evolved numerous morphological, behavioral, and physiological differences from their surface conspecifics (6,7). These fish can be efficiently reared in laboratories, allowing for the application of transgenic and gene-editing approaches (8). There is a rapidly growing focus on genomic data in these systems that compare cave and surface populations. Current genomic data includes fully assembled genomes for surface and cave populations, population genetic resequencing, and transcriptomic data across different contexts (9,10). The development of a database that compiles the growing number of genomics data across different contexts would provide a valuable resource for accessing and analyzing this information.

The Shiny package in R offers a method to produce powerful community web resources that can go far beyond traditional repositories of data (11). Shiny databases enable researchers to incorporate the statistical analysis and data visualization

capabilities of the R programming language into a reactive database that also functions as a community data repository. The combination of these tools allows users to sift through vast amounts of data, enabling novel discoveries (12). The generation of a Shiny database for comparative models of evolution could combine data across populations and studies. The flexibility of these systems and intrinsic analysis capabilities allows for direct comparisons of genetic data from disparate sources. Here, we generated a Shiny database, CaveCrawler, which combines population genetics and transcriptomic data from multiple Mexican tetra populations and leverages Gene Ontology (GO) term information to enable unique biological inferences from cross-study patterns. We demonstrate that the analysis features of this program can identify genes that are implicated in evolutionary processes across populations of *A. mexicanus*, using different methodologies, and in different studies.

Methods

The CaveCrawler database acts as a repository for transcription, Gene Ontology (GO), population genetics, and annotated genome data acquired from different studies in *A. mexicanus*, including those using reference genomes for surface and Pachón cavefish (9,13). With a highly accessible web interface, CaveCrawler enables researchers to search for data on genes-of-interest, find genes whose transcriptional levels match defined criteria, find genes which fit desired population genetics parameters, and also identify genes associated with cellular components, molecular functions, and biological processes.

CaveCrawler modules

The CaveCrawler framework utilizes a bifurcated design with an underlying data repository and a collection of user interface modules (Figure 1). The databases currently offers five user modules: Home, Gene Search, Transcription, Population Genetics, and GO Term Info. Each of these modules is designed to draw on different elements of the underlying data repository. This bifurcated design facilitates simple updates to the repository which then are immediately populated into changes in the functionality and results produced by the modules that draw on the updated repositories. Similarly, new

modules can be added at any time to take advantage of new types of analyses users desire or new data types included in the repository. The home module houses general information about *A. mexicanus* and about CaveCrawler's functionality, as well brief instructions for contributing data.

The Gene Search module enables the user to search for data associated with genes-of-interest and also to identify genes associated with GO terms-of-interest. In this module, the user inputs a single gene stable ID, a single GO term, or a comma-separated list of genes. The module outputs a downloadable table describing all genes associated with the inputs and the positional, transcription, and population genetics data associated with each of the genes. The output also indicates whether a statistic or piece of transcriptional data is not present for each gene-of-interest. Therefore, this module concatenates data from disparate sources into a single analysis output, enabling the user to efficiently search for existing data and identify experiments which have yet to be conducted on their genes-of-interest.

The Transcription module enables the user to identify genes which differ in transcription level between groups. Here, the user first inputs the groups they would like to compare. The user may either compare an experimental group to a control group or compare one morph to another morph. The user then specifies whether they would like to see genes which are up or downregulated in the first group compared to the second and the percent change in transcription level between groups. The module then produces a downloadable output table of genes fitting the specified transcription patterns.

The Population Genetics module enables the user to access population genomics statistics, such as π, Tajima's D, d_{XY}, and F_{ST}. This module has two options for accessing population genomics data. In the first option, the user provides GO terms and the module outputs and visualizes the statistical values of all genes associated with those GO terms. The second approach enables the user to search for transcriptional or genomic values associated with defined across different analyses.

In the GO term search function of the Population Genetics module, the user inputs GO information, statistics-of-interest, and populations-of-interest. For the GO information, the user can input either a single GO ID, a comma-separated list of GO IDs, or a phrase associated with the target GO term. The module outputs a downloadable table describing all values of the population-specific statistics-of-interest for the genes associated with the indicated GO term(s). If any of the statistics-of-interest require pairwise comparisons between populations, the module will output pairwise statistics for each possible pairing of input populations. On this submodule, the user may also input a statistic and a scaffold and CaveCrawler will plot the statistical values of each GO-term-associated gene which falls on that scaffold. The GO term function of the Population Genetics module thus enables the user to access and visualize population genomics statistics for a GO term of interest.

The outlier function of the Population Genetics module consists of two approaches for pulling outlier genes from combined datasets. One approach enables the user to identify a specified number of genes which have the most extreme values for an indicated statistic, while the other approach enables the user to identify all genes whose statistic value falls above or below a specified threshold value. In the gene number approach, the user must specify the number of genes and must specify whether they would like to see the top or bottom quantile. CaveCrawler then outputs a table describing the specified number of genes with the most extreme values for the statistic-of-interest. In the statistical threshold approach, the user specifies a threshold statistical value and specifies whether they would like to see genes above or below this value. CaveCrawler outputs both a table and a distribution plot describing the genes which fall above or below this threshold.

Both outlier approaches require the user input a statistic-of-interest and population(s)-of-interest. If the statistic-of-interest is a one-population statistic, such as π or Tajima's D, both approaches will report outlier statistical values for all input populations. If the input statistic is a pairwise statistic, such as F_{ST} or d_{XY}, both approaches will report outlier statistical values for all possible pairs of populations-of-interest. If a statistic value has yet to be collected for a population or population pair, CaveCrawler will output a warning

about that statistic. Thus, the outlier function of the Population Genetics module enables users to not only identify outliers for a statistic-of-interest but also to identify populations for which a statistic-of-interest has yet to be collected.

The GO Term Info module enables users to access descriptions of GO IDs. This function helps users identify GO IDs they should search for in the Population Genetics module and helps them make sense of transcription and outlier queries. On this module, the user may input a single GO ID, comma-separated list of GO IDs, or a phrase-of-interest, such as “sleep”. CaveCrawler searches data from the official Gene Ontology databank, outputting descriptions of all input GO IDs or GO IDs relevant to the input phrase. In addition, CaveCrawler reports all GO IDs which occur hierarchically beneath these IDs. The GO Term Info module thus enables researchers to investigate the broader biological impact of transcription and diversity data relevant to their genes-of-interest.

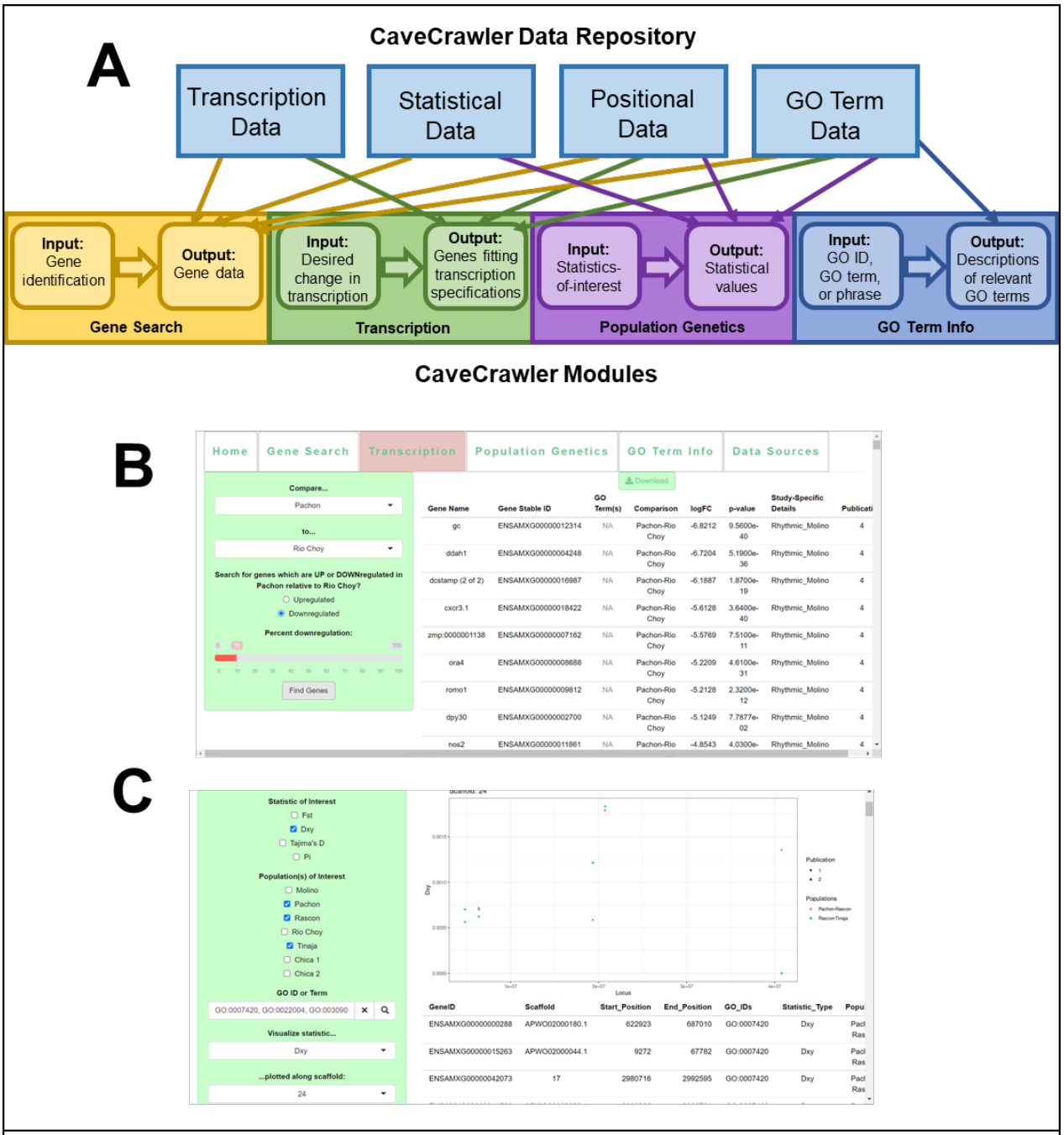

***Figure 1*. Design and web interface for CaveCrawler**

A) The repository and module framework for the CaveCrawler model organism genomics database. Lines show the connections between different types of data stored in the repository and the user modules that draw on each data type. B) Example of the Transcription module with the results of searching for top 10% of genes that are downregulated in Pachón relative to Río Choy surface fish C) Example of the Population Genetics module with the results of searching for the Pachón-Rascon surface fish and Rascon-Tinaja d_{XY} values of genes associated with brain development GO IDs and visualizing these values on Scaffold 24

169 ***The data repository***

170 CaveCrawler pools data from multiple publications and authors can request that their
171 own data be integrated into CaveCrawler's repository. As of publication, CaveCrawler's
172 data bank includes transcriptional datasets (14,15), population genetics datasets (10,15),
173 GO data from UniProt and the Gene Ontology Consortium (16-18), and genome
174 architecture data from Ensembl Genome Browser, release 104 (19).
175
176 CaveCrawler's Transcription and Gene Search modules currently draw upon datasets
177 that describe genes whose transcription levels changed significantly in response to sleep
178 deprivation in *A. mexicanus* (20). This dataset describes the log fold-change (logFC) and
179 p-values for each of these genes in each *A. mexicanus* morph where the change in
180 transcription was significant compared to controls of the same morph (20). As described
181 in the Transcription module section of this paper, CaveCrawler can also access
182 transcription data for genes whose transcription is significantly different between morphs
183 (14). The Transcription module has enough flexibility that new transcriptional data can be
184 integrated. Thus, CaveCrawler could be used to analyze transcriptional changes in
185 response to any experimental condition and between any two morphs of *A. mexicanus.*
186
187 CaveCrawler's Population Genetics and Gene Search modules currently integrate
188 data from two studies describing signatures of selection in *A. mexicanus* (10,15). One of
189 these studies calculated π and Tajima's D values for the Pachon, Tinaja, Molino, Río
190 Choy, and Rascon populations, as well as F_{ST} and d_{XY} values for each population pair
191 (10). The other study describes d_{XY} values of all genes in two populations of the Chica
192 morph, Pachon and Rascon, and Tinaja and Rascon (15). As with the Transcription
193 module, the Population and Gene Search modules have enough flexibility that new data
194 can be integrated.
195
196 The Gene Search, Transcription, and Population Genetics modules currently draw
197 upon positional data obtained from Ensembl (19). The genome assembly used in the
198 current version is *A. mexicanus* 2.0, the most up-to-date genome assembly for this

199 species (9). All of CaveCrawler's modules utilize GO term information from UniProt and
200 from the Gene Ontology Consortium (16-18).
201
202 Though CaveCrawler already integrates data from numerous disparate sources,
203 enabling powerful cross-study comparisons of genetic data, CaveCrawler's data
204 repository is not static. The CaveCrawler website includes instructions for data
205 submission and the power and insights possible with this resource will grow as the
206 repository of data on which draws grows. CaveCrawler's data repository will be updated
207 annually in July.
208
209 **Results**
210 The CaveCrawler analysis suite consists of multiple tools for comparing datasets that
211 allow for identification of genetic differences between populations of *A. mexicanus*. These
212 tools have a wide range of applications, including rapid candidate gene identification and
213 inference of population-level variation. Here, we present an example of how CaveCrawler
214 can be used to answer biological questions.
215
216 ***Rapid Identification of Candidate Genes for Empirical Studies***
217 Since CaveCrawler enables simultaneous cross-analysis of multiple studies,
218 researchers can use CaveCrawler to find genes which are outliers for both transcription
219 and population genetics statistics in a matter of minutes. These genes can then be
220 analyzed in downstream studies, such as GO term analyses, to make biological
221 inferences. Here, we identified genes which are transcriptionally dysregulated between
222 cave and Río Choy morphs, then performed a GO term analysis to determine the
223 biological function and cellular components with which these genes are associated.
224 These genes could be used as candidates for future empirical studies, such as
225 knockdown or knockout studies.
226
227 To examine genes that are both transcriptionally upregulated and harbor markers of
228 selection, we first used CaveCrawler's Population Genetics module to identify the F_{ST}
229 values of all genes whose F_{ST} values were published in a recent population inference

paper in the Mexican tetra (see 10). Then, we used the 'Gene Search' module to identify the transcription data for each of the 1140 genes identified by the Population Genetics module (see Supplemental Table 1). By using the Gene Search module, we found that previous studies had measured the between-morph logFC values for 267 of the 1140 genes for which we had F_{ST} values. Pairwise F_{ST} measures how dissimilar a DNA sequence is between two groups relative to diversity within the groups, and logFC is the log fold change in mRNA transcription between two groups (14,21). Of the genes for which both F_{ST} and logFC had been calculated by previous studies, there were 72 for which F_{ST} outlier status had been determined by a previous study (10). Gene names, logFC, transcription p-values, and F_{ST} values for all 72 genes are available in the supplemental materials.

For each cave- Río Choy surface comparison, we then identified the genes which were both significantly differentially expressed for circadian regulation (logFC p-value < 0.05) between Río Choy and the corresponding cave population and were identified by a previous study to be F_{ST} outliers for the same population pairing (10). These genes, which were both transcriptional and F_{ST} outliers, will henceforth be referred to as double outliers. We found one gene which was a double outlier in all three cave-Río Choy pairings (Table 1; Figure 2), one which was a double outlier for both Pachón-Río Choy and for Tinaja-Río Choy (Table 1; Figure 2A and 2C), one which was a double outlier for Molino-Río Choy only (Table 1; Figure 2B), one which was a double outlier for Tinaja-Río Choy only (Table 1; Figure 2C).

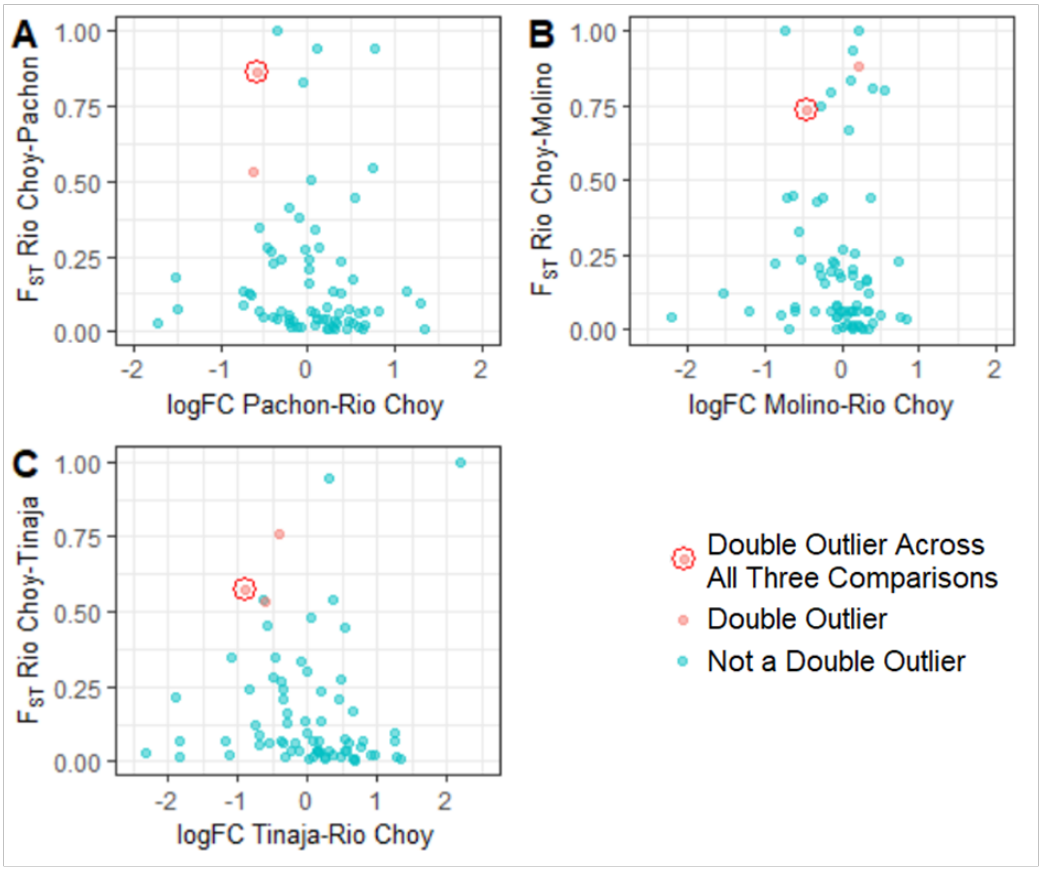

Figure 2: **Overlap between F_{ST} values and logFC values across multiple studies.** Plots of cave-specific F_{ST} vs. logFC values for all 72 genes which CaveCrawler found to have F_{ST} values, logFC values, and F_{ST} outlier designations. Double-outliers for the cave-Río Choy comparison indicated by the axes are colored in red, while the gene (*arpin*) which was a double-outlier across in all 3 cave-Río Choy comparisons is encircled in red. A) Pachón vs. Río Choy B) Molino vs. Río Choy C) Tinaja vs. Río Choy.

We performed a GO term analysis on *arpin to* identify any biological process, molecular function, or cellular component associated with this double outlier. We found *arpin* to be associated with the biological process GO ID GO:0051126 and the cellular component GO IDs GO:0016021 and GO:0030027*,* which correspond to “negative regulation of actin nucleation”, “integral component of membrane”, and “lamellipodium”, respectively. To calculate the likelihood of sampling an *A. mexicanus* gene associated with GO:0051126 by chance, we performed a Monte Carlo simulation for 1000000 iterations and calculate an empirical p-value of 2.8e-05. We performed another Monte Carlo to find the

likelihood of sampling GO:0016021 and GO:0030027 by chance, obtaining an empirical p-value of 4.4e-05. Thus, we used CaveCrawler to rapidly discover that genes that harbor markers of selection and are transcriptionally in cave populations across the circadian cycle.

As shown by this example, the CaveCrawler analysis suite can be used for a variety of investigations in the Mexican tetra. CaveCrawler can in minutes combine statistics from multiple studies and leverage GO terms to make novel inferences about evolutionary forces acting within a population.

Discussion

Here, we describe a modular analysis suite for *A. mexicanus.* We have included a set of the genomics and transcriptional data that has been previously published. In addition to these studies, transcriptional analysis across developmental timepoints, as well as single cell analysis of hypothalamus has been collected. These data sets, and others collected in the future can be added to this analysis suite. These data, in combination with assembled genomes for surface fish and Pachón cavefish provide a platform for gene discovery in this system. In addition, the modularity of this system allows it to be readily adapted for new data types or genomic analyses. We then demonstrated that this analysis suite can be used to combine data from disparate sources to discover novel patterns in the Mexican tetra genome.

As proof of principle, we performed an analysis for genes that contained markers of selection and transcriptional dysregulation across the circadian cycle. This analysis identified four genes that were significantly different. These genes represent strong candidate for functional regulators of evolved differences in circadian behavior that have been widely studied in *A. mexicanus* and other species of cavefish (14,22-25). The gene *arpin*, a negative regulator of *actin* is of particular interest because it is identified as harboring markers of selection and transcriptional dysregulation across all three cavefish populations. Actin dynamics have been implicated as targets of circadian regulation for a number of processes including wound healing, immune function and neural plasticity (26-28). Therefore, it is possible that multiple populations of cavefish have converged on

changes in actin regulation that account for loss of behavioral and transcriptional rhythms (14,24).

Shiny has been widely applied to develop a range of public databases that offer interactive data visualization and access (12,29,30). However, to our knowledge, this is the first use of Shiny to create a public genomic database and analysis tool for any model organisms. Traditionally these resources which are key to supporting model organism communities have come with considerable cost in the form of computer programmers and hosting services (31,32). Perhaps one of the most valuable contributions that CaveCrawler can make is as a flexible framework that can be adopted by any model organism community. We have made the underlying code for this project publicly available under the GPL license. All source code and example datasets are available in the GitHub repository: https://github.com/AnnabelPerry/AstyanaxShinyApp.

In *A. mexicanus*, like many other models of evolution, studies identifying quantitative trait loci (QTL) have provided a basis for a growing genetic toolkit in *A. mexicanus* can be used for functional genomics experiments guided CaveCrawler (7,33). For example, transgenesis, CRISPR-based transgenesis, and morpholinos have all been applied for functional validation of gene function (34-37). In addition, CRISPR-based screening approaches have been developed in zebrafish that allow for high throughput functional assessment of developmental and behavioral traits. This analysis suite will provide methodology for identifying genes for functional analysis.

Gene Name	Comparison	Double Outlier	F_{ST}	logFC	p-value for logFC
si:dkeyp-84f3.5	Pachón vs. Río Choy	No	0.277988	0.139042	0.055824
	Molino vs. Río Choy	Yes	0.882812	0.225267	0.003513
	Tinaja vs. Río Choy	No	0.300349	-0.01207	0.8625
socs6b	Pachón vs. Río Choy	No	0.826635	-0.04327	0.83292
	Molino vs. Río Choy	No	0.836271	0.107438	0.54766
	Tinaja vs. Río Choy	Yes	0.756493	-0.40817	0.01563
cyp26a1	Pachón vs. Río Choy	Yes	0.53226	-0.61497	0.007395
	Molino vs. Río Choy	No	0.792368	-0.14028	0.55483
	Tinaja vs. Río Choy	Yes	0.536471	-0.59503	0.009671
arpin	Pachón vs. Río Choy	Yes	0.861159	-0.58471	0.000358
	Molino vs. Río Choy	Yes	0.734267	-0.46234	0.000122
	Tinaja vs. Río Choy	Yes	0.576169	-0.88163	1.33E-10

***Table 1*: Genes identified as outliers for F_{ST} and transcriptional regulation over the circadian cycle between surface fish and three different cavefish populations.**
F_{ST} and logFC values for all genes which were found to be outliers for both F_{ST} and logFC in at least one cave-Río Choy comparison

Acknowledgements

This work was supported by an NIH NIGMS R35GM138098 to HB, NIH R01 1R01GM127872 to ACK, and SEM, NIH R21 NS122166 to ACK, and the Texas A&M University College of Science Undergraduate Research Opportunities Program to ARP

Reference

1. Harris, T.W., Arnaboldi, V., Cain, S., Chan, J., Chen, W.J., Cho, J., Davis, P., Gao, S., Grove, C.A., Kishore, R. *et al.* (2020) WormBase: a modern Model Organism Information Resource. *Nucleic Acids Res*, **48**, D762-D767.
2. Howe, D.G., Ramachandran, S., Bradford, Y.M., Fashena, D., Toro, S., Eagle, A., Frazer, K., Kalita, P., Mani, P., Martin, R. *et al.* (2021) The Zebrafish Information Network: major gene page and home page updates. *Nucleic Acids Res*, **49**, D1058-D1064.
3. Larkin, A., Marygold, S.J., Antonazzo, G., Attrill, H., Dos Santos, G., Garapati, P.V., Goodman, J.L., Gramates, L.S., Millburn, G., Strelets, V.B. *et al.* (2021) FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Res*, **49**, D899-D907.
4. Juntti, S. (2019) The future of gene-guided neuroscience research in non-traditional model organisms. *Brain, behavior and evolution*, **93**, 108-121.
5. McGaugh, S.E., Kowalko, J.E., Duboué, E., Lewis, P., Franz-Odendaal, T.A., Rohner, N., Gross, J.B. and Keene, A.C. (2020). Wiley Online Library.
6. Gross, J.B. (2012) The complex origin of Astyanax cavefish. *BMC evolutionary biology*, **12**, 1-12.
7. Jeffery, W.R. (2020) Astyanax surface and cave fish morphs. *EvoDevo*, **11**, 1-10.
8. Klaassen, H., Wang, Y., Adamski, K., Rohner, N. and Kowalko, J.E. (2018) CRISPR mutagenesis confirms the role of oca2 in melanin pigmentation in Astyanax mexicanus. *Developmental Biology*, **441**, 313-318.
9. Warren, W.C., Boggs, T.E., Borowsky, R., Carlson, B.M., Ferrufino, E., Gross, J.B., Hillier, L., Hu, Z., Keene, A.C. and Kenzior, A. (2021) A chromosome-level genome of Astyanax mexicanus surface fish for comparing population-specific genetic differences contributing to trait evolution. *Nature communications*, **12**, 1-12.
10. Herman, A., Brandvain, Y., Weagley, J., Jeffery, W.R., Keene, A.C., Kono, T.J., Bilandžija, H., Borowsky, R., Espinasa, L. and O'Quin, K. (2018) The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, Astyanax mexicanus. *Molecular ecology*, **27**, 4397-4416.
11. Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. and Borges, B. (2021) shiny: Web Application Framework for R. R package version 1.6.0.
12. Blackmon, H. and Demuth, J.P. (2015) Coleoptera karyotype database. *Coleopt. Bull*, **69**, 174-175.
13. McGaugh, S.E., Gross, J.B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., Hinaux, H., Jeffery, W.R., Keene, A. and Ma, L. (2014) The cavefish genome reveals candidate genes for eye loss. *Nature communications*, **5**, 1-10.
14. Mack, K.L., Jaggard, J.B., Persons, J.L., Roback, E.Y., Passow, C.N., Stanhope, B.A., Ferrufino, E., Tsuchiya, D., Smith, S.E. and Slaughter, B.D. (2021) Repeated evolution of circadian clock dysregulation in cavefish populations. *PLoS genetics*, **17**, e1009642.
15. Moran, R.L., Jaggard, J.B., Roback, E.Y., Rohner, N., Kowalko, J.E., Ornelas-Garcia, P., McGaugh, S.E. and Keene, A.C. (2021) Hybridization underlies localized trait evolution in cavefish. *bioRxiv*.

16. The UniProt Consortium. (2020) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**, D480-D489.
17. The Gene Ontology Consortium. (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, **49**, D325-D334.
18. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S. and Eppig, J.T. (2000) Gene ontology: tool for the unification of biology. *Nature genetics*, **25**, 25-29.
19. Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Research*, **49**, D884-D891.
20. McGaugh, S.E., Passow, C.N., Jaggard, J.B., Stahl, B.A. and Keene, A.C. (2020) Unique transcriptional signatures of sleep loss across independently evolved cavefish populations. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **334**, 497-510.
21. Charlesworth, B. (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular biology and evolution*, **15**, 538-543.
22. Teyke, T. and Schaerer, S. (1994) BLIND MEXICAN CAVE FISH (ASTYANAX HUBBSI) RESPOND TO MOVING VISUAL STIMULI. *J Exp Biol*, **188**, 89-101.
23. Moran, D., Softley, R. and Warrant, E.J. (2014) Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism. *PLoS One*, **9**, e107877.
24. Beale, A., Guibal, C., Tamai, T.K., Klotz, L., Cowen, S., Peyric, E., Reynoso, V.H., Yamamoto, Y. and Whitmore, D. (2013) Circadian rhythms in Mexican blind cavefish Astyanax mexicanus in the lab and in the field. *Nat Commun*, **4**, 2769.
25. Ceinos, R.M., Frigato, E., Pagano, C., Fröhlich, N., Negrini, P., Cavallari, N., Vallone, D., Fuselli, S., Bertolucci, C. and Foulkes, N.S. (2018) Mutations in blind cavefish target the light-regulated circadian clock gene, period 2. *Sci Rep*, **8**, 8754.
26. Hoyle, N.P., Seinkmane, E., Putker, M., Feeney, K.A., Krogager, T.P., Chesham, J.E., Bray, L.K., Thomas, J.M., Dunn, K., Blaikley, J. *et al.* (2017) Circadian actin dynamics drive rhythmic fibroblast mobilization during wound healing. *Sci Transl Med*, **9**.
27. Kitchen, G.B., Cunningham, P.S., Poolman, T.M., Iqbal, M., Maidstone, R., Baxter, M., Bagnall, J., Begley, N., Saer, B., Hussell, T. *et al.* (2020) The clock gene Bmal1 inhibits macrophage motility, phagocytosis, and impairs defense against pneumonia. *Proc Natl Acad Sci U S A*, **117**, 1543-1551.
28. Petsakou, A., Sapsis, T.P. and Blau, J. (2015) Circadian Rhythms in Rho1 Activity Regulate Neuronal Plasticity and Network Hierarchy. *Cell*, **162**, 823-835.
29. Manchanda, N., Portwood, J.L., Woodhouse, M.R., Seetharam, A.S., Lawrence-Dill, C.J., Andorf, C.M. and Hufford, M.B. (2020) GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. *BMC genomics*, **21**, 1-9.
30. Consortium, T.o.S. (2014) Tree of Sex: A database of sexual systems. *Scientific Data*, **1**.
31. Oliver, S.G., Lock, A., Harris, M.A., Nurse, P. and Wood, V. (2016) Model organism databases: essential resources that need the support of both funders and users. *BMC biology*, **14**, 1-6.

32. Bellen, H.J., Hubbard, E., Lehmann, R., Madhani, H.D., Solnica-Krezel, L. and Southard-Smith, E.M. (2021) Model organism databases are in jeopardy. *Development*, **148**, dev200193.
33. Casane, D. and Rétaux, S. (2016) Evolutionary Genetics of the Cavefish Astyanax mexicanus. *Adv Genet*, **95**, 117-159.
34. Elipot, Y., Legendre, L., Père, S., Sohm, F. and Rétaux, S. (2014) Astyanax transgenesis and husbandry: how cavefish enters the laboratory. *Zebrafish*, **11**, 291-299.
35. Jaggard, J.B., Stahl, B.A., Lloyd, E., Prober, D.A., Duboue, E.R. and Keene, A.C. (2018) Hypocretin underlies the evolution of sleep loss in the Mexican cavefish. *Elife*, **7**.
36. Stahl, B.A., Jaggard, J.B., Chin, J.S.R., Kowalko, J.E., Keene, A.C. and Duboué, E.R. (2019) Manipulation of Gene Function in Mexican Cavefish. *J Vis Exp*.
37. Stahl, B.A., Peuß, R., McDole, B., Kenzior, A., Jaggard, J.B., Gaudenz, K., Krishnan, J., McGaugh, S.E., Duboue, E.R., Keene, A.C. *et al.* (2019) Stable transgenesis in Astyanax mexicanus using the Tol2 transposase system. *Dev Dyn*, **248**, 679-687.