bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Type of the Paper (Article)

Developmental transcriptomic analysis of the cave-dwelling crustacean, *Asellus aquaticus*

Joshua B. Gross [2], Dennis A. Sun [3**] , Brian M. Carlson [4], Sivan Brodo-Abo [1] , and Meredith E. Protas [1*]**

[1] Department of Natural Sciences and Mathematics, Dominican University of California, San Rafael, California 94901
[2] Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio 45221
[3] Department of Molecular and Cellular Biology, University of California at Berkeley, Berkeley, California 94720
[4] Department of Biology, College of Wooster, Wooster, Ohio 44691

** equal contribution
* Correspondence: meredith.protas@dominican.edu

Received: date; Accepted: date; Published: date

Abstract: Cave animals are a fascinating group of species often demonstrating characteristics including reduced eyes and pigmentation, metabolic efficiency, and enhanced sensory systems. *Asellus aquaticus*, an isopod crustacean, is an emerging model for cave biology. Cave and surface forms of this species differ in many characteristics, including eye size, pigmentation and antennal length. Existing resources for this species include a linkage map, mapped regions responsible for eye and pigmentation traits, sequenced adult transcriptomes, and comparative embryological descriptions of the surface and cave forms. Our ultimate goal is to identify genes and mutations responsible for the differences between the cave and surface forms. To advance this goal, we decided to use a transcriptomic approach. Because many of these changes first appear during embryonic development, we sequenced embryonic transcriptomes of cave, surface, and hybrid individuals at the stage when eyes and pigment become evident in the surface form. We generated a cave, a surface, a hybrid, and an integrated transcriptome to identify differentially expressed genes in the cave and surface forms. Additionally, we identified genes with allele-specific expression in hybrid individuals. These embryonic transcriptomes are an important resource to assist in our ultimate goal of determining the genetic underpinnings of the divergence between the cave and surface forms.

Keywords: regressive evolution, *de novo* transcriptome, differential expression, troglomorphy, cave

1. Introduction

Cave animals are fascinating organisms that frequently share a common suite of characteristics, including reduced eyes, reduced pigmentation, metabolic differences, and enhanced sensory systems. Questions that have long fascinated cave biologists include how and why these characteristics have evolved, and, whether the same underlying mechanisms mediate trait loss between different cave populations and different cave species.

Historically, it has been challenging to understand how and why cave characteristics have evolved, due to difficulties with rearing cave organisms in captivity, and a lack of contemporary experimental resources (e.g., genomic, genetic, and functional molecular tools) for most cave species. In recent years, however, there have been vast expansions of, and improvements in, resources and tools for emerging model organisms. Obtaining genomic information is now possible for most systems, and the complete genomic sequence is

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

available for a limited number of cave dwelling species [1,2]. In addition, many studies have involved transcriptome sequencing projects for cave dwellers such as crayfish, salamanders, amphipods, isopods, and fish [3-8]. For the vast majority of these projects, adult samples have been utilized due to the challenge of obtaining embryonic samples of natural, cave dwelling species. However, for several cave species, many trait differences are established early in embryonic development, underscoring the importance (and value) of analyzing gene expression differences across embryonic development.

The star among the cave species is *Astyanax mexicanus*, where it is possible to work with embryos and obtain embryonic samples, as well as perform genetic analyses (reviewed in [9,10]). Both adult and embryonic transcriptomes have been generated, as well as a draft genome sequence for cave and surface morphs [2,11,12]. Additionally, contemporary genomic tools, such as gene editing, provide the ability to functionally analyze candidate genes discovered through transcriptome sequencing [13-15]. Because of the wealth of data provided by these emerging resources, historical questions impacting on the evolution of cave animals can now be addressed (reviewed in [16,17]).

Despite the great deal of information provided by decades of research in *A. mexicanus*, additional studies from other cave organisms are necessary to understand the convergence of regressive loss across animals that inhabit the cave biome. Specifically, the mechanisms that mediate regressive loss in *A. mexicanus* may differ from those mechanisms operating in other cave-adapted species. Thus, it is important to develop other species in a similar way to *A. mexicanus* in order to widen our perspective, and to gain a broader understanding of how cave evolution occurs across diverse taxa.

Unfortunately, not every cave animal is amenable to develop as a model in the same way as *A. mexicanus*. There are many considerations, foremost of which is the ability to raise and breed a species in the lab. This feature greatly reduces the number of cave-adapted species for which genetic and developmental studies are feasible. Generally, only limited information is available regarding the life cycle and ecology of cave dwelling organisms, rendering it difficult to recapitulate the natural environment in the lab. Another important feature for these investigations is an extant surface-dwelling form capable of interbreeding with cave morphs. Owing to the divergence times between cave and surface morphs, the ability to produce viable hybrid offspring is very unusual among studied cave organisms.

Asellus aquaticus is a freshwater crustacean that has two morphs: a cave and surface form, both of which can be raised in the lab [18]. Interbreeding between cave and surface forms was first documented in the 1940s [19]. Much of the historical work on *Asellus aquaticus* has included comparative morphology between the surface and cave forms, and population genetic analyses of several cave and surface populations throughout Europe [20-25]. Recently, a classical genetics approach has been made possible by multiple crossing strategies to create F1, F2, and backcross pedigrees between cave and surface populations. These studies have resulted in production of a linkage map, insight into the genetic architecture of this species, and identification of genomic regions associated with different cave-associated phenotypes [26-28] .

Though advances have been made in genomic mapping alongside the development of genetic resources, the identity of genes responsible for these trait differences between cave

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and surface forms remain unknown. A powerful approach to identifying genetic differences between cave-and surface-dwelling forms is comparative transcriptomics. Transcriptomes have been characterized for multiple cave populations, including the Pivka channel of the Planina cave and the Molnár János cave [6,29]. Though these studies have been useful in generating genetic resources, the causative genes mediating differences between cave and surface populations have not been established. Part of the issue, as discussed above, is that adult samples are not the most appropriate, as many different characteristics between cave and surface individuals are established during embryonic development [28,30]. For example, eye loss and pigment loss are established at the end of embryogenesis. To investigate the genetic pathways responsible for eye and pigment loss, the most appropriate samples to sequence would be those obtained at this timepoint in embryonic development.

To address this gap in knowledge, we generated *de novo* embryonic transcriptomes from one cave and one surface population, as well as from hybrid individuals. We hypothesized that many genes would be differentially expressed between cave and surface forms, including those involved in neurogenesis, pigment development, eye development, and metabolism. Furthermore, we expected that a subset of these differentially expressed genes would also show allele-specific expression, suggesting that regulatory mutations result in altered transcriptional abundance for those genes.

2. Materials and Methods

Animals

Animals were collected from Rakov Škocjan location (surface) and the Rak Channel of the Planina cave population (cave) (Figure 1A). Animals were reared in water, lighting, and food conditions as previously described [26-28]. Surface animals were raised in tanks with around 10 individuals per tank. Similarly, cave animals were raised in tanks with around 10 individuals per tank. Hybrid crosses were generated by mating a single cave male to a single surface female. When a female with embryos was observed in any of the above tanks, the females were monitored until the embryos were around 70% of the way through development. They were then removed from the female using a clove oil solution of 20µl in 50 ml of fresh water as previously described [28]. Embryos were kept in a small dish with commercial spring water (Crystal Geyser) until they reached 90% of embryonic development, when both pigmentation and incipient ommatidia were present in the surface, but not cave, embryos (Figure 1B, C) [28].

RNA extraction, library preparation and sequencing

An entire brood was used for a single sample which ranged from 25 - 89 embryos. Embryos were extracted in 200µl of Trizol and mechanically disrupted using an Eppendorf pestle. Samples were sent to the Functional Genomics Lab, Vincent J. Coates Genomics Sequencing Laboratory, California Institute for Quantitative Biosciences (QB3) University of California, Berkeley. Total RNA was extracted using the Trizol protocol. PolyA selection was performed and library preparation was performed using the low input protocol of the Nugen kit. Sequencing was performed using 150 bp paired end reads on both the Illumina Hiseq 4000 and the Hiseq 2500 sequencing machines.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

de novo *transcriptome assembly and annotation*

A total of 36 sequencing files were processed for transcriptome assembly and annotation. We evaluated three *Asellus* cave embryonic samples (MPD1, MPD5, MPD6), three surface embryonic samples (MPD2, MPD3, MPD8) and three hybrid embryonic samples (MPD4, MPD7, MPD9) subjected to pair-read sequencing and processed in duplicate (total = 36 files). To achieve the most accurate mapping for downstream RNA-seq studies, we built morphotype-specific transcriptomes using SeqMan NGen (DNAStar). Initial mapping utilized default parameters (mer size: 21, minimum match percentage: 80%, cluster size: 100,000) which resulted in numerous incompletely assembled contigs. To increase the average transcript lengths of our assemblies, we tested a variety of parameters. We found the optimal results when we adjusted the mer size (19), and increased the minimum match percentage (to 97%), and maximum cluster size (to 300,000). This approach provided the longest mean transcript lengths (surface = 1061bp, cave = 1069bp, hybrids = 952bp), as well as the most assembled transcripts > 1kb in length (surface = 49,233; cave = 51,822; hybrids = 52,390; Table 1). We reasoned that the longest transcripts represented the best individual transcript assemblies, and therefore proceeded to annotate those assembled transcripts that were 1000bp or longer.

All annotations were carried out using Blast2GO (v.5.2.5) running Java v.1.8.0_144. To capture the most comprehensive information, we performed two rounds of BLAST-associated annotations for each of three transcriptomes - one using the *Tribolium castaneum* genome as a reference, and one using the SwissProt database (Table 2). In brief, we submitted a fasta-formatted file containing all *de novo*-assembled sequences to Blast2GO, specified our database of interest, and proceeded through all default annotation steps. We implemented a script to remove all annotated transcripts associated with ribosomal or mitochondrial sequences, which ranged between 734 – 1066 sequences with an identified blast hit. For all three transcriptomes (surface, cave, and hybrids), we obtained comparable results for both databases, however the *Tribolium castaneum* reference provided the largest number of successful annotations.

RNA-sequencing and expression analyses

Once annotation was completed, we performed RNA-sequencing analyses using ArrayStar (v.13; DNAStar, Madison). For each assembly, we performed duplicate RNA-seq analyses for all transcriptome references (i.e., *Tribolium castaneum* and SwissProt), and retrieved very similar results. Accordingly, developmental sequencing reads were aligned from all three morphotypes (cave, surface and hybrids), and normalized using RPKM to control for variation in sequencing depth and transcript length. The resulting dataset included a measure of linear total RPKM, which provided a single metric of expression that could be compared across datasets. We subsequently used this metric to calculate fold change differences between groups (e.g., cave versus surface).

We tested the robustness of our assemblies by using several reference files, which allowed us to compare between assemblies to evaluate the consistency of calculated expression. This annotation process periodically yielded >1 blast hit to a single, orthologous

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reference transcript. Therefore, we averaged the RPKM values for all assembly contigs with a blast identity to the same reference transcript in order to estimate the most accurate expression level for a given dataset. This calculation enabled us to correct for multiple blast hits to the same reference, however it may have inadvertently collapsed the expression for different isoforms (or paralogues) into a single transcript. We acknowledge this potential confounding issue, however we note that different isoforms are catalogued in both the *Tribolium* reference and SwissProt. In cases where multiple transcripts for a single *Tribolium* gene were present in our dataset, we excluded these results from our analysis, as it was not possible to determine whether these represented genuinely distinct *Asellus aquaticus* transcripts, rather than sequences representing paralogues or alleles harboring significant changes. Therefore, this project could not assess the possibility of *Asellus aquaticus*-specific isoforms or paralogous genes, a caveat that will need to be addressed in future genome sequencing projects. Finally, given the inaccessibility of fresh tissues (with which to extract RNA for quantitative PCR validation), we implemented a variety of strict filters to maximize the likelihood that our reported differentially-expressed genes were valid.

Allele-specific expression using ASE-TIGAR

To assess allele-specific expression of differentially expressed genes, pairs of transcripts were identified across cave and surface transcriptomes if they had the same *Tribolium castaneum* Uniprot ID. For a given pair of alleles, transcripts were manually trimmed to be similar in length, based on sequence identity (Figure 2B; Supplementary File 1). We then used the ASE-TIGAR software [31] to generate transcript abundances for each allele. The software was supplied a single FASTA file containing both trimmed alleles from the cave and surface transcriptomes, as well as paired-end reads from the MPD4, MPD7, and MPD9 hybrid embryo transcriptomes. The output of this software was a file containing the expected number of fragments mapped by ASE-TIGAR, an FPKM value, and a THETA value, which was the estimated transcript abundance. We used this THETA value as our metric of expression for each allele. Given that the list of genes we selected for allele-specific expression analysis could be biased towards genes that might show allele-specific expression, we determined that it was important to have a statistically rigorous approach to identifying genes with true allele-specific expression differences. An ideal null distribution for hypothesis testing in this scenario would be the distribution of all log fold change values for all pairs of genes. However, generating such a dataset was neither practical nor computationally feasible. Instead, we chose to simulate a null distribution that represented the intra-allele variance using the THETA values calculated for each allele in each replicate (MPD4, MPD7, MPD9). This null distribution would convolve noise arising from technical differences (batch effects, sequencing errors, etc.) and biological differences (gene expression variability between samples, gene expression noise, etc.). We generated an intra-allele null distribution by comparing inter-replicate log fold changes for all replicates within a given allele, e.g. gene X, surface allele replicate 1 vs. gene X, surface allele replicate 2, etc., using a custom Python script. Genes that had THETA = 0 in one or more replicates of one or more alleles were filtered out of the analysis. We then compared the distributions of

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

intra-allele variations for surface alleles and cave alleles using a two-sample Kolmogorov-Smirnov (K-S) test and found that the two distributions were indistinguishable (K-S statistic = 0.0289, p-value = 0.9643). We merged the surface and cave allele null distributions and used this total distribution as a null distribution for assessing significance, also using a two-sample Kolmogorov-Smirov (K-S) test. For each pair of alleles, we generated a distribution of log fold changes by comparing each replicate of one allele to each replicate of the other allele, for a total of 9 values per pair of alleles. We used the two-sample K-S test implemented in the Pandas Python package to generate a K-S statistic and a p-value, and then performed the Benjamini-Hochberg (B-H) multiple hypothesis testing correction procedure to that p-value using a Scipy.stats Python package and $\alpha = 0.05$. Genes for which significant log fold change differences were observed based on this B-H corrected p-value were called as genes with true allele-specific expression differences.

GO term analysis and enrichment studies

Enrichment analysis was performed of genes that showed significant allele-specific expression using the online gene ontology tool [32,33]. Biological process and molecular function were selected and the genome searched was *Tribolium castaneum.*

Data deposition

All sequences analyzed in this report have been provisionally submitted to the National Center for Biotechnology Information, Sequencing Reads Archive (BioProject ID:xxxxxxx).

3. Results

3.1 Characterization of surface, cave and hybrid transcriptomes

Following the optimization of our assembly parameters, we retrieved highly similar results for all three of our assembled transcriptomes (Table 1). The total number of reads that were assessed for each transcriptome was very similar between surface morphs (~364M), cave morphs (~361M), and hybrids (~394M). The total number of assembled reads for surface (~155M), cave (~164M) and hybrids (~132M) were similarly comparable, although a higher proportion of assembled reads were utilized in cave morphs (45.5%), compared to surface morphs (42.4%) and hybrids (33.5%). The reduced proportion of assembled reads used in the hybrid transcriptome assembly may reflect the sequence divergence between cave and surface morphs. Overall, we feel the unassembled reads that were not incorporated into each assembly is attributable to the relatively high amount of repeat sequence present in the sampled libraries (data not shown).

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Comparison of transcriptome assemblies of *Asellus* cave morphs, surface morphs and hybrid individuals.

	Surface morphs	Cave morphs	Hybrids
Sequence Read Summary			
Total Assembled Reads	155039720	164487662	132336702
Total Unassembled Reads	83386227	109047509	84947885
Total Reads Excluded by Sampling	126373422	87592389	176811792
Total Number of Reads	364799369	361127560	394096379
Transcript Summary			
Total number of Transcripts	113432	119569	143962
Average Length of Assembled Transcripts	1061	1069	952
Assembled Transcripts >1kb	49,233	51,822	52,390
Assembly Time	50.7 hours	54.7 hours	54.2 hours

All assemblies were completed in roughly the same amount of time (50.4 hr – 54.2 hr), and yielded comparable numbers of transcripts (surface = 113K; cave = 119K; hybrid = 143K), or comparable average lengths (surface = 1061bp; cave = 1069bp; hybrid = 952 bp). Our goal, however, was to annotate the best-characterized transcripts in each dataset. We reasoned that the longest transcripts represented the best individual transcript assemblies, and therefore proceeded to annotate those assembled transcripts that were 1000bp or longer. This value was similar across all three assemblies: surface = 49,233; cave = 51,822; hybrids = 52,390 (Table 2).

Table 2. Annotation results against two reference databases for *Asellus* cave morphs, surface morphs and hybrid *de novo* transcriptomes.

	Surface morphs	Cave morphs	Hybrids
Tribolium Genome Database			
Total Number of Transcripts	113,432	119,569	143,962
Assembled Transcripts >1kb	49,233	51,822	52,390
No BLAST hits	28,648	30,340	30,709
Ribosomal sequences	518	749	712
Mitochondrial sequences	880	734	973
Total number of annotated sequences	19,187	19,999	19,996

	Surface morphs	Cave morphs	Hybrids
SwissProt Database			
Total Number of Transcripts	113,432	119,569	143,962
Assembled Transcripts >1kb	49,233	51,822	52,390
No BLAST hits	29,918	31,928	32,157
Ribosomal sequences	603	624	839
Mitochondrial sequences	986	992	1066
Total number of annotated sequences	17,726	18,278	18,328

Using these assemblies as a starting point, we subjected each transcriptome to comprehensive annotation using Blast2GO (Methods). This BLAST-based approach was performed against the *Tribolium castaneum* genome and SwissProt database, in order to compare the quality of each database. We chose these databases because *Tribolium castaneum* is an arthropod with a comprehensive genome database, and the SwissProt database is an open-access and manually annotated resource of protein sequence and functional information. Overall, we found that the average percentage of failed BLAST hits was higher when we used the SwissProt database (mean = 61.3%) compared to the *Tribolium castaneum* database (mean = 58.4%). Consequently, our final transcriptome size was larger when we annotated against the *Tribolium* (mean = 19,727 transcripts) compared to the SwissProt database (mean = 18,110 transcripts). In sum, our results indicated that the

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tribolium castaneum database provided better results (Table 2), and therefore our downstream analyses utilized these annotated transcriptomes.

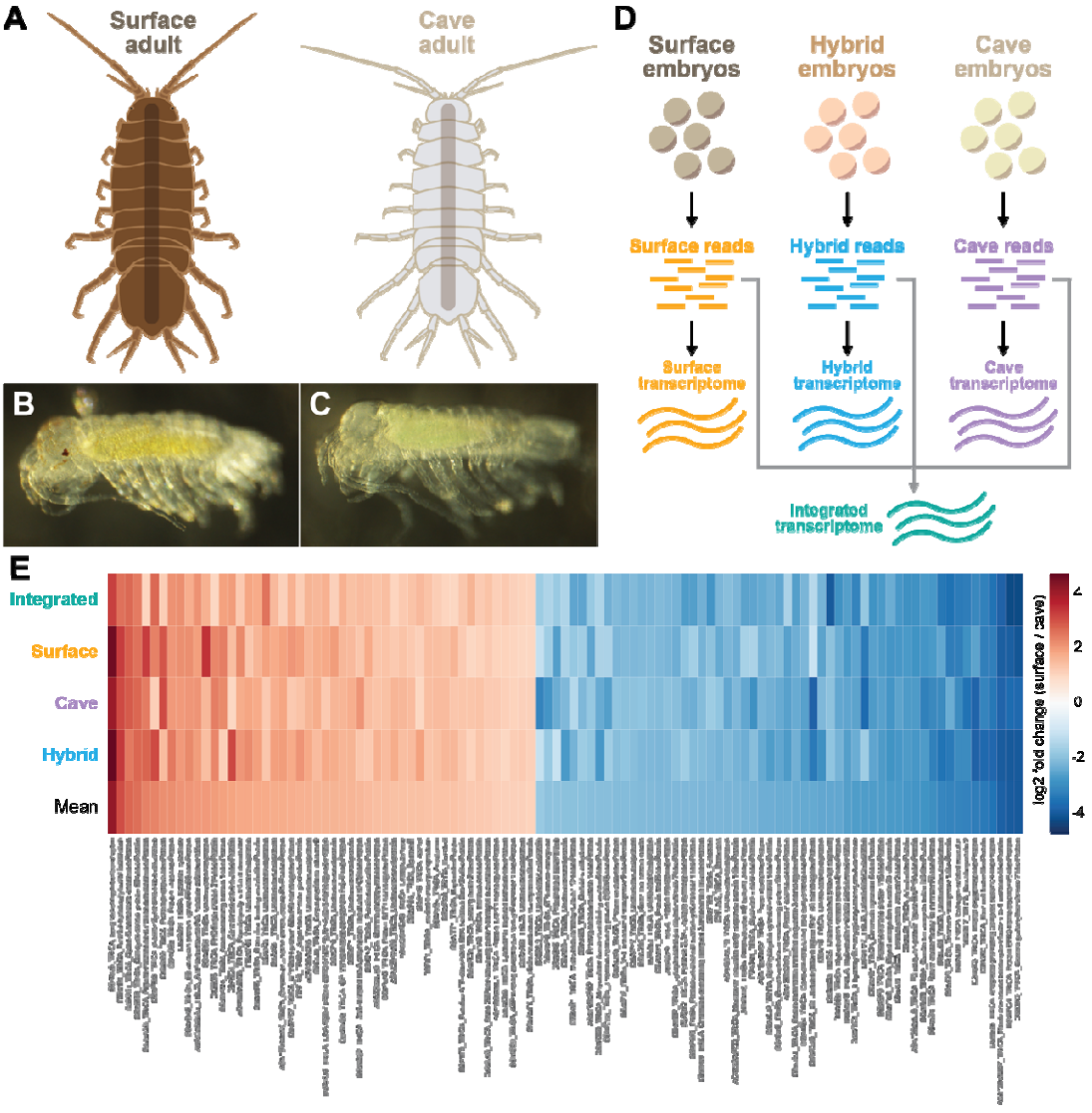
3.2 Differential RNA-seq analysis between cave and surface morphs

We mapped the cave and surface reads separately to each of the four different transcriptomes: cave, surface, hybrid, and integrated transcriptomes. We selected all genes that had at least a two-fold change in the same direction (increased or decreased expression) between cave and surface in all four experiments and had a standard deviation of less than or equal to 8. Then we selected the top 50 genes that were underexpressed in the cave form and the top 50 genes that were overexpressed in the cave form to analyze further (Figure 1E; Supplementary File 2).

Several of the genes that were underexpressed in the cave made biological sense, as they are involved in eye or pigment function such as *long-wavelength sensitive opsin*, *cell cycle control protein 50A-like*, *membrane-bound transcription factor site 1 protease-like protein*, *scarlet-like protein*, *protein pygopus-like*, and *atonal*. Genes that were overexpressed in the cave form include those involved in metabolism, such as *solute carrier family 35 member F6-like protein*, *gamma-glutamyltransferase 7-like protein*, and *inositol oxygenase-like protein*. Also overexpressed in the cave samples was *annulin-like protein* which is expressed in stripes in each limb bud segment [34] and could be a candidate for differential antennal characteristics in the cave form.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1: Top 50 overexpressed genes and top 50 underexpressed genes in the cave samples as compared to the surface samples. A. Illustrations of a surface adult and a cave adult. Cave adults lack pigmentation and eyes, and have longer limbs. B. Surface embryo that has gone through 90% of embryonic development. C. Cave embryo that has gone through 90% of embryonic development. D. Four different transcriptomes were generated, one from the cave embryonic samples, one from surface embryonic samples, one from hybrid embryonic samples, and one from all embryonic samples (referred to as the integrated transcriptome). E. Heatmap showing the top 50 downregulated genes in the cave form (various shades of red) or top 50 upregulated genes in the cave form (various shades of blue). All genes shown had the same direction of fold change and a standard deviation of less than eight across all 4 analyses. Uniprot ID and gene name from the *Tribolium castaneum* genome is shown.



bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

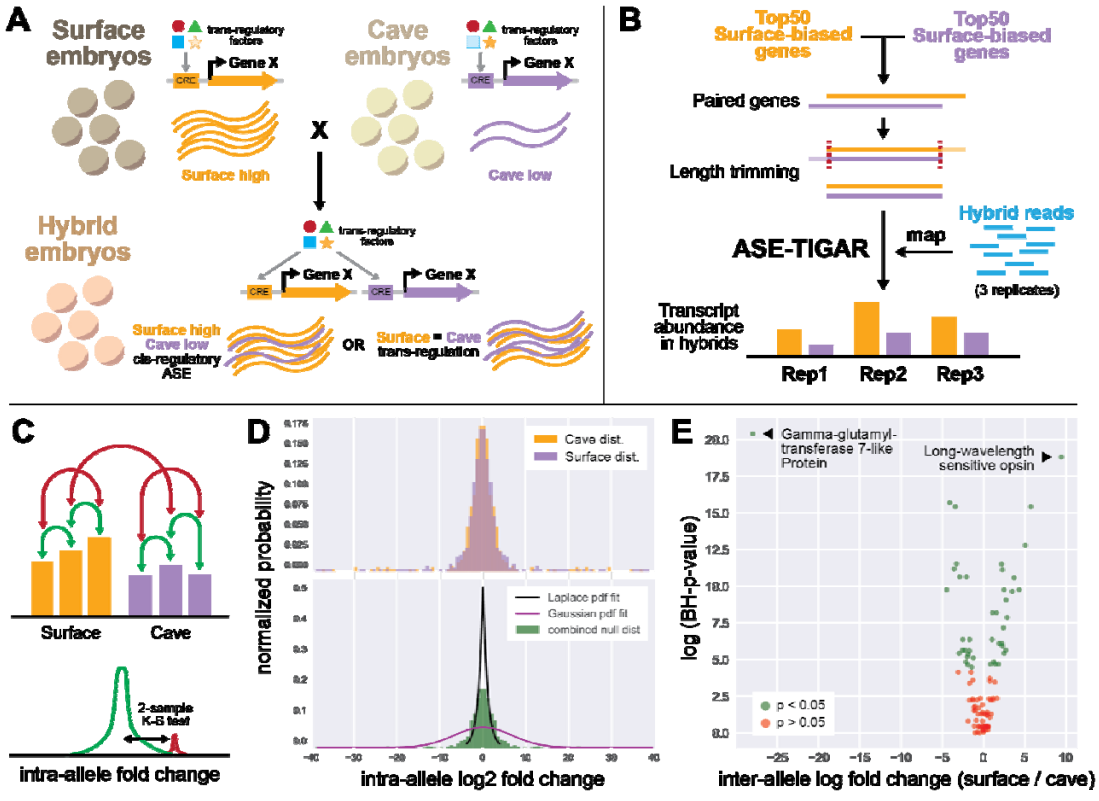
3.3 Allele-specific expression analysis reveals pervasive cis-regulatory mechanisms for gene expression differences between cave and surface populations

Genes that display differential expression between populations may arrive at this difference through both *cis*- and *trans*-regulatory mechanisms. In *cis*-regulatory changes to gene expression, a change to the DNA sequence either within a gene or in regulatory elements thereof is responsible for an observed expression difference between populations (Figure 2A). When *trans*-regulatory factors change gene expression, the regulatory sequence of a gene may not change, but instead, a change to the expression of a trans-regulatory factor (an activator, repressor, etc.) between populations drives the difference in expression of a downstream gene. By examining the expression of alleles of a given gene in hybrid organisms, one can determine mechanisms of gene expression difference, whether they be *cis*-regulatory, *trans*-regulatory, or a combination of both. In hybrid animals, *trans*-regulatory effects are normalized across alleles, as both alleles existing in the same nucleus are subjected to the same input by activators and repressors. As such, when expression differences in alleles are observed in hybrids, one possible explanation is that *cis*-regulatory changes contribute to differential expression between populations (Figure 2A). Allele-specific differences can also come about due to parent-of-origin effects, though our experimental design should prioritize genes with *cis*-regulatory changes rather than parent-of-origin effects (see Discussion).

We wanted to examine the mechanism of differential gene expression for the genes we identified as differentially expressed between cave and surface populations. To do this, we performed allele-specific expression (ASE) analysis using the ASE-TIGAR software package [31]. This software, given a FASTA file containing both isoforms of a gene and FASTQ reads from hybrid animals, generates transcript abundance estimates for each allele (Figure 2B). We identified pairs of alleles for the most differentially expressed genes and generated a log fold change value for the usage of surface vs. cave alleles in hybrid animals (Figure 2). We then used a two-sample Kolmogorov-Smirnov test with a Benjamini-Hochberg multiple hypothesis testing correction to call significance of observed ASE, using intra-allele log fold change as our null distribution (see Methods; Figure 2C, D). Overall, genes with significant ASE tended to have larger log2 fold change between the two alleles (Figure 2E; Supplementary File 2).

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2: Experimental design of allele-specific expression analysis. A. A hypothetical example of gene X which is differentially expressed between cave versus surface individuals and also shows allele-specific expression with lower expression of the cave allele. In this case, the underlying mechanism may be a result of differences between *trans*-regulatory factors between populations, or in the *cis*-regulatory sequence. If *cis*-regulatory mechanisms dominate, then we expect to see similar surface-biased expression in hybrids. However, if *trans*-regulatory mechanisms dominate, then the equalized *trans*-regulatory environment in hybrids will result in no allele-specific expression. B. Pipeline of ASE analysis. The top 50 differentially expressed genes, in both directions, present in both the cave and surface transcriptomes were selected, paired genes were trimmed to the same length and hybrid reads were mapped to the trimmed cave and surface versions of each gene. C. Intra-allele log2 fold change was calculated by comparing transcript abundance between replicates of a given allele (green arrows). This null distribution (green curve) was then compared to distributions of inter-allele log2 fold change (red arrows, red curve) using a two-sample Kolmogorov-Smirnov test. D. Intra-allele log2 fold change distributions for cave and surface populations. The top panel shows each allele separately, overlapped. These distributions were determined to be indistinguishable (K-S test, see Methods). The bottom panel shows the combined distribution (green), and a Laplace fit (black line) and Gaussian fit (magenta line) to the distribution. The combined null distribution does not neatly fit either a Laplace or Gaussian distribution, validating that the two-sample Kolmogorov-Smirnov test is appropriate, as it does not assume that either distribution is parametric. E. The analysis identified 43 genes that had significant allele-specific expression (green) and 54 genes that did not have significant allele-specific expression.



bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Many of the genes we identified as highly differentially expressed (DE) between individuals of different populations also appeared to show ASE between alleles in hybrid animals (Figure 2E, Figure 3A and B; Supplementary File 2). For example, the *long-wavelength sensitive opsin* gene was found to be about 4-fold (mean log2 fold change across transcriptomes) underexpressed in cave than surface animals (mean = 4.15 (log2 scale), SEM = 0.33), and was the most surface-biased gene by DE analysis. In hybrid animals containing one surface and one cave allele, we observed that the same gene showed a 10-fold (mean log2 fold change) difference between alleles (mean = 10.497, SEM = 2.79). This suggests that *cis*-regulatory changes contribute to differences in *long-wavelength sensitive opsin* expression between populations.

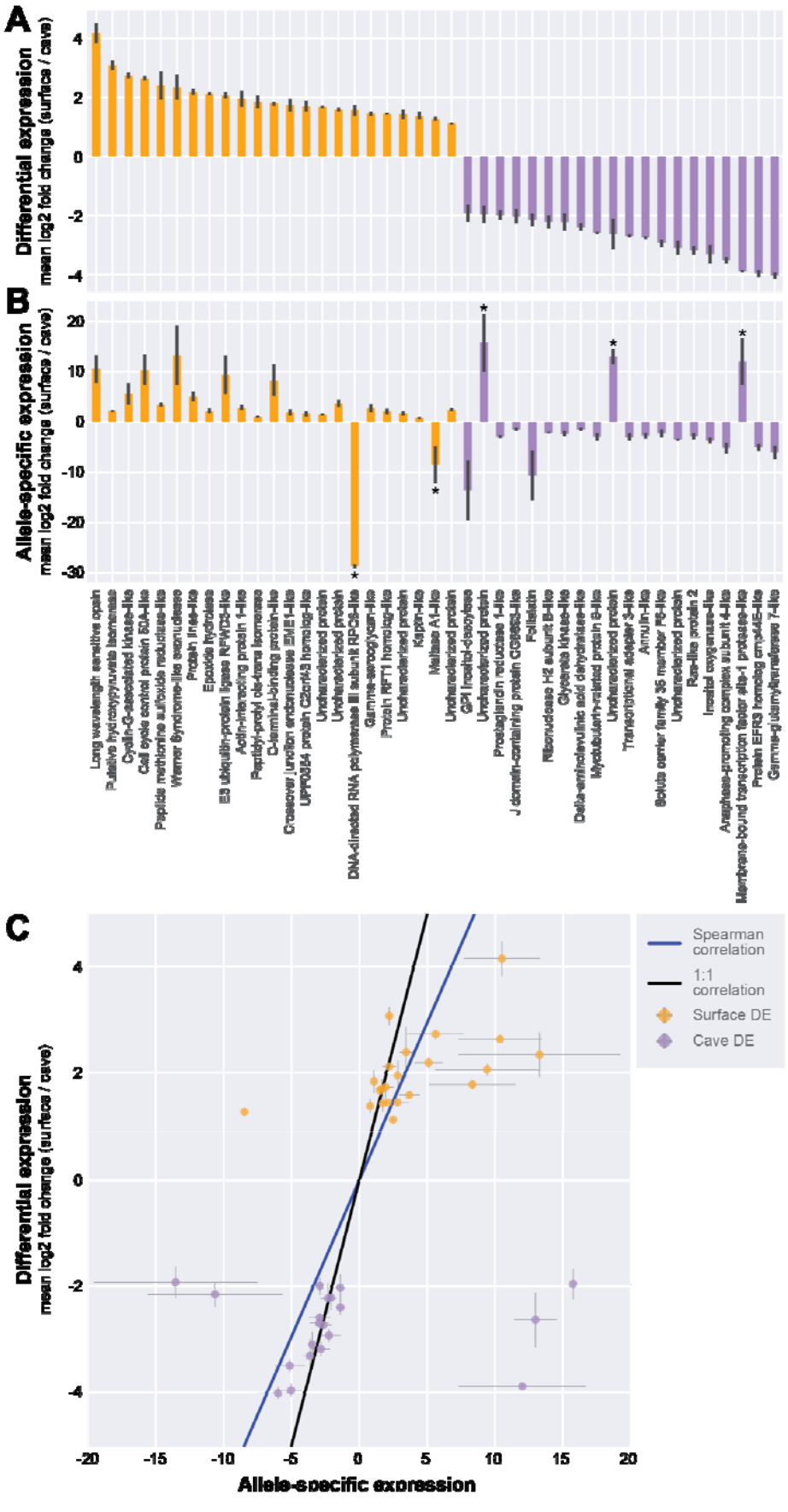
By examining all genes with significant ASE, we observed that most of the genes had some cis-regulatory component to their change in expression between populations. We inferred this result because genes that showed DE in favor of surface animals, on the whole, tended to also show ASE in favor of the surface allele (21 out of 23 genes, Figure 3A and 3B). Meanwhile, genes that showed DE in favor of cave animals also tended to have ASE in favor of the cave allele (17 out of 20 genes, Figure 3A and 3B). For five genes (Figure 3B, marked with asterisks), we observed significant ASE that showed a strong bias in the opposite direction from what we expected from the DE analysis. For example, *DNA-directed RNA polymerase III subunit RPC8-like Protein* and *Maltase A1-like Protein* were found to be more highly expressed in surface animals, but by ASE the cave allele appeared to be more expressed. Such results can be explained through models of competing *cis*-by-*trans* effects.

A GO enrichment analysis was performed of the genes that showed significant allele-specific expression. Enrichment was seen in the GO molecular function complete category of catalytic activity (FDR 3.69×10^{-2}) and the GO biological process category of metabolic process (FDR 2.23E-2). No significant results were seen when separating the genes showing allele-specific expression into genes that showed a bias towards the cave allele or a bias towards the surface allele.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3: Allele-specific expression suggests *cis*-regulatory contribution to population difference. A. For the subset of genes that showed significant allele-specific expression, mean log2 fold change comparing surface and cave differential expression. B. Mean log2 fold change comparing surface allele and cave allele expression within the hybrid, in the same order as in A. C. A Spearman correlation test indicates that DE and ASE analyses are significantly correlated (correlation = 0.5897, p-value = 3.1659e-10).

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4. Discussion

Candidate genes

Typical features of cave animals include loss of eyes, loss of pigment, differences in metabolism, and enhanced sensory structures. Specifically in *Asellus aquaticus*, the cave form can show loss of eyes, loss of pigment, and increased appendage length [23,24,35]. Less is known about metabolic and behavioral differences between the cave and surface populations, but a recent study showed that acetylcholinesterase and glutathionine S transferase had lower activity in cave individuals as compared to the surface individuals, supporting the idea that the cave form has lower metabolic and locomotor activity [36]. In addition, shelter-seeking behavior has been shown to be different between some cave and surface populations [37]. Overall, we expected to find differential expression and allele-specific expression in genes involved in eye development, pigmentation, appendage development, and metabolism. As expected, some of the differentially expressed genes that we found to be differentially expressed have been shown to play a role in phototransduction, photoreceptor development, and/ or eye development such as *atonal, long-wavelength sensitive opsin, cell cycle control protein 50A-like, membrane-bound transcription factor site 1 protease like protein, Protein EFR3 homolog cmp44E-like protein, pygopus-like protein*, and *domeless*. Furthermore, a subset of the above, (*long-wavelength sensitive opsin, cell cycle control protein 50A-like, membrane-bound transciption factor site 1 protease like protein*, and *EFR3 homolog cmp44E-like protein)* also showed allele-specific expression indicating that *cis*-regulatory changes may be responsible for the differential expression of those genes. Fewer genes with known involvement in pigmentation were observed. *Scarlet*, a gene involved in pigment transport [38], was overexpressed in the surface form; however, *scarlet* was not shown to have allele-specific expression and therefore is unlikely to have a *cis*-regulatory change. *Annulin-like protein* was another gene of interest which was overexpressed in the cave form as compared to the surface form and had higher cave allele expression in the hybrids. Interestingly, this gene is expressed in grasshoppers in stripes along the forming limb segments and could be a candidate for appendage length changes in *A. aquaticus* [39]. Another gene of interest which had showed higher ASE for the cave allele was *Myotubularin related protein 9 like protein* (*MTMR9*). Polymorphisms in this gene have been shown to be associated with obesity and glucose tolerance in GWAS studies in humans [40,41]. It is unknown whether *Asellus aquaticus* has any adaptive behaviors or features regarding food acquisition in the cave environment but, studies in the cavefish *Astyanax mexicanus* have shown that some cave populations are insulin resistant and able to binge eat [42,43].

Another interesting gene that showed both expression differences between populations and allele-specific differences is *gamma-glutamyl transferase 7-like protein* (*GGT7*). Elevated GGT is commonly seen in individuals with non-alcoholic fatty liver disease [44]. Interestingly, one of the cave populations of the cavefish *Astyanax mexicanus* develops fatty livers when exposed to high nutrient conditions [42]. Little is known about fat storage in *Asellus aquaticus*, but future studies can examine whether cave and surface forms of *Asellus aquaticus* differ in fat storage, insulin resistance, and starvation resistance similar to cave and surface populations of *Astyanax mexicanus*.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

447

448 *Involvement of regulatory mutation versus coding mutation in evolution of cave traits*

449 When working with species with limited genomic and genetic resources, most studies
450 that discover the causative genes for particular phenotypes involve coding mutations. This
451 may be due to ascertainment bias, as coding mutations are much easier to identify than
452 *cis*-regulatory mutations, which could be in much larger (and uncharacterized) regions of the
453 genome. Furthermore, *cis*-regulatory changes can be more difficult to test functionally than
454 coding mutations. Because of these challenges, most of the mutations and genes identified as
455 causative for cave-related traits in the model system of *Astyanax mexicanus* have been
456 coding mutations [42,45-50] though there are some exceptions [51]. Allele-specific
457 expression studies in hybrids are a powerful way of identifying *cis*-regulatory differences.
458 Here, we have identified many genes with allele-specific expression, some of which likely
459 have *cis*-regulatory changes, as inferred through a positive correlation between
460 allele-specific expression and differential expression. Studies have indicated that much of
461 evolutionary change occurs via *cis*-regulatory mutations (reviewed in [52]) and therefore, the
462 establishment of techniques to identify such changes in species that have limited genomic
463 and genetic resources is crucial for identifying the genetic/genomic substrate of evolutionary
464 change.

465

466 Cis- *versus* Trans- *regulation*

467 Another major question in evolutionary biology regards whether *cis*- or *trans*-regulatory
468 changes dominate in driving evolutionary change. In *trans*-regulatory changes, modifications
469 to the expression or function of *trans*-regulatory factors, such as transcription factors, have a
470 cascading effect on the expression of many other downstream target genes, driving
471 evolutionary changes. *Cis*-regulatory changes, on the other hand, are more restricted, tend to
472 occur in regulatory regions, and affect the expression of a particular gene. *Trans*-regulatory
473 change might be expected to cause more drastic and pleiotropic effects, whereas
474 *cis*-regulatory change would be less likely to have pleiotropic consequences (reviewed in
475 [53]). Previous studies comparing species and interspecific hybrids have shown input of
476 both *trans*-and *cis*-regulatory change (reviewed in [53]).

477 In our study, we have observed evidence of changes in both *cis*- and *trans*- regulation
478 between *Asellus aquaticus* populations. An example of a likely *cis*-regulatory change is in the
479 case of the *long-wavelength sensitive opsin* gene, for which cave samples showed lower
480 expression than surface samples. In hybrid samples, the cave allele also showed significantly
481 lower expression as compared to the surface allele; the shared directionality of the DE and
482 ASE results for this gene suggests that cis-regulatory effects are responsible for expression
483 differences between the populations. On the other hand, the *scarlet* gene appears to be an
484 example of a *trans*-regulatory change in our dataset. Here, though the cave samples showed
485 lower expression as compared to the surface samples, in hybrid samples, the cave allele was
486 not significantly reduced in expression compared to the surface allele. When both *scarlet*
487 alleles were placed in an identical *trans*-regulatory environment, the alleles expressed at
488 indistinguishable levels, suggesting that differences in a *trans*-regulatory factor between the
489 populations is responsible for expression differences. However, we cannot exclude

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cis-regulation for this gene as it is possible that certain genes show allele-specific expression only in specific tissues and sequencing transcriptomes of entire bodies dilutes out any tissue specific allele-specific expression [54]. In addition, we observed cases wherein cave samples showed lower expression compared to surface samples, but in hybrids the cave allele had higher expression. This might indicate both *trans* and *cis* modes of regulation, which may be evidence for compensatory mutations (reviewed in [53]). However, a recent study indicated that cases of compensatory *cis*-and *trans*-regulation are often overestimated as a result of correlated errors that occur when estimating ASE [55]. Our data was not amenable to the analysis presented in this paper, due to the methods we used to estimate ASE. The interplay between *cis*-and *trans*-regulation can ideally be examined by comparing the fold change of DE versus ASE. If the fold change of ASE is equal to the fold change of DE, *cis*-regulation likely explains the differential expression fully [53]. If the fold change of ASE is less than the fold change of DE, a combination of *cis*-regulation and *trans*-regulation likely explains the differential expression. Our DE and ASE analyses used different measures of transcript abundance and therefore the fold changes of each are not directly comparable. Future analyses with greater sample sizes and different measures of transcript abundance may enable finer examination of the differences in ASE and DE for other genes.

Parent-of-origin effects versus cis*-regulation*

Allele-specific expression in hybrid organisms can result from *cis*-regulatory change or because of parent-of-origin effects, in which the maternal and paternal copies of the gene are expressed differently, as has been observed in different organisms, including mammals, insects, and plants [56]. If our study had examined allele-specific expression genome-wide in an unbiased fashion, it is likely that some of the genes showing allele-specific expression would be due to parent-of-origin effects. However, our study has examined allele-specific expression only in those genes that showed differential expression in cave versus surface samples. We would not necessarily expect genes that have parent-of-origin effects to also show differential expression in cave versus surface samples. Therefore, though we cannot exclude parent-of-origin effects in genes that are differentially expressed in cave versus surface individuals, it is likely that many of the genes in which we see allele-specific expression have *cis*-regulatory changes, rather than parent-of-origin effects. Future studies can eliminate potential parent-of-origin effects by generating hybrid samples from both cave female x surface male and surface female x cave male matings. As the former crosses are considerably more difficult to generate, our study was restricted to samples from the latter type of cross. Future investigations may tease apart the genes that are truly expressed as a result of cis-regulatory changes versus those with parent-of-origin effects, once it is more tractable to generate crosses with cave female and surface male animals.

Comparison to adult Asellus *transcriptome*

There are two previously published transcriptomes from *Asellus aquaticus*, both on mostly adult samples [29,57,58] . The first transcriptome utilized Roche/454 sequencing technology and was more limited in terms of actual sequence generated, though some surface embryonic samples were sequenced [6,57]. This transcriptome was generated from the Pivka channel of

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the Planina Cave in Slovenia. More recently, a transcriptome was generated from Hungarian populations of *Asellus aquaticus*, including the Molnár Janós cave population. This study found that genes involved in phototransduction were still expressed in this cave population [29]. The authors found two expressed opsins, and neither seemed to have drastic coding changes. Consequently, they hypothesized that decreased expression of the opsins could be behind vision loss in this population. This idea is supported by our study, which uses a different cave population. Specifically, we found both differential and allele-specific expression in *long-wavelength sensitive opsin*. Our studies have expanded the transcriptomic resources for this species by generating a transcriptome for an additional cave population, the Rak channel of the Planina cave. This is a useful cave population to examine as comparative embryology, as well as genetic mapping studies, have both been performed for this cave population [26,28]. In addition, this is the first study generating an embryonic transcriptome of a cave population of *Asellus aquaticus* and examining differential and allele-specific expression between cave and surface individuals, giving us a window into the developmental mechanisms resulting in population-specific differences.

Comparison to other cave dwelling animal transcriptomes

Transcriptomes of many cave dwelling organisms have now been sequenced. Examples include other populations of *Asellus aquaticus*, *Gammarus minus* (an amphipod crustacean) [59] , *Niphargus hrabei* (another amphipod crustacean; [29]), cave crayfishes [60,61], *Poecilia mexicana* [5], *Sinocycloheilus* species [4,62], multiple species of cave beetles [63,64], multiple isopod species [65] and *Astyanax mexicanus* [11,12]. Transcriptome studies of these cave animals often look to see whether genes involved in phototransduction are still expressed and whether there are any obvious mutations in genes involved with vision. The majority of the transcriptomes described above are from adult samples, owing to difficulties with breeding or otherwise obtaining embryonic samples. However, embryonic samples have been examined in *Astyanax mexicanus* [6,12].

One approach that has been lacking in studies of cave transcriptomes is using hybrid transcriptomes to evaluate allele-specific expression. A previous study in *Asellus aquaticus* examined allele-specific expression in a limited number of genes from a single adult hybrid sample [6]. In *Astyanax mexicanus*, allele-specific expression has not yet been documented from hybrid individuals. In most other cave-dwelling animals it is not possible to examine allele-specific expression because it requires both a cave and surface form, and they must be capable of interbreeding. However, here we show that where this approach is possible, it is a powerful way to investigate genes that might have *cis*-regulatory mutations. In the future, this approach can be applied to other species that have surface and cave forms, even those that might not have fertile hybrids or viable hybrids (as long as the hybrids can start development). Potential species to examine include *Poecilia mexicana* and *Gammarus minus*.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Future steps

Here, we examined comparative expression and allele-specific expression in whole bodies of groups of individuals at a particular developmental timepoint. In the future, we aim to expand our analysis to additional timepoints and potentially specific tissues, as these two factors are known to influence both comparative and allele-specific expression [54]. Additionally, now that methods are established to investigate differential expression and allele-specific expression in embryonic samples of cave versus surface morphs of *Asellus aquaticus*, one next step is to expand the analysis to other cave populations. One of the advantages of working with this species is the number of populations that are thought to be independently evolved [25,66]. By examining gene expression and allele expression differences in these different cave populations, it should be possible to better understand how these cave specific traits have evolved, and determine if they have evolved similarly or differently in the independently evolved populations. Furthermore, now that a number of candidates with putative *cis*-regulatory changes have been identified, we can investigate them by placing them to the linkage map to determine if they coincide with mapped regions responsible for eye and pigment variation. Also, future work developing functional methods in *Asellus aquaticus*, such as genome editing and gene expression visualization, will enable testing of these genes to validate whether they are causative for associated cave related traits.

Supplementary Materials: Supplementary File 1, Supplementary File 2

Author Contributions: J.B.G and D.A.S designed the experiments, carried out the experiments, performed the analyses, and wrote the manuscript. B.M.C carried out the experiments, performed analyses, and wrote the manuscript. S.B.A carried out the experiments and reviewed the manuscript. M.E.P designed the experiments, carried out the experiments, and wrote the manuscript.

Funding: MEP is supported by funding from the US National Institutes of Health (NEI R15 EY029499-01A1). DAS is supported by an NSF GRFP Fellowship (Division of Graduate Education, Grant/Award Number: NSF GRFP 2016230010). JBG is supported by funding from the US National Institute of Health (NIDCR R01-DE025033) and the US National Science Foundation (DEB-1457630).

Acknowledgments: Thanks to Peter Trontelj, Žiga Fišer, Teo Delić, and Gregor Bračko for animal collection and/or advice throughout the project. Thanks to Hafasa Mojaddidi, Hana Whitehorn, and Emily Klein for technical assistance. Thanks to Vincent J. Coates Genomics Sequencing Laboratory, California Institute for Quantitative Biosciences (QB3) University of California, Berkeley for the transcriptome sequencing services. In addition, we would like to thank Nicholas Lammers and Nicholas Everetts for their helpful advice on the statistical methods performed in this manuscript. Thanks also to Rachel Brem and Tal Ronnen Oron for advice about allele-specific expression analysis.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] Yang, J.; Chen, X.; Bai, J.; Fang, D.; Qiu, Y.; Jiang, W.; Yuan, H.; Bian, C.; Lu, J.; He, S. *et al.* The Sinocyclocheilus Cavefish Genome Provides Insights into Cave Adaptation. BMC Biol. **2016**, *14*, 4.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[2] McGaugh, S.E.; Gross, J.B.; Aken, B.; Blin, M.; Borowsky, R.; Chalopin, D.; Hinaux, H.; Jeffery, W.R.; Keene, A.; Ma, L. *et al.* The Cavefish Genome Reveals Candidate Genes for Eye Loss. Nat. Commun. **2014**, *5*, 5307.

[3] Carlini, D.B.; Satish, S.; Fong, D.W. Parallel Reduction in Expression, but no Loss of Functional Constraint, in Two Opsin Paralogs within Cave Populations of Gammarus Minus (Crustacea: Amphipoda). BMC Evol. Biol. **2013**, *13*, 8-89.

[4] Meng, F.; Braasch, I.; Phillips, J.B.; Lin, X.; Titus, T.; Zhang, C.; Postlethwait, J.H. Evolution of the Eye Transcriptome Under Constant Darkness in Sinocyclocheilus Cavefish. Mol. Biol. Evol. **2013**, *30*, 1527-1543.

[5] Passow, C.N.; Brown, A.P.; Arias-Rodriguez, L.; Yee, M.C.; Sockell, A.; Schartl, M.; Warren, W.C.; Bustamante, C.; Kelley, J.L.; Tobler, M. Complexities of Gene Expression Patterns in Natural Populations of an Extremophile Fish (Poecilia Mexicana, Poeciliidae). Mol. Ecol. **2017**.

[6] Stahl, B.A.; Gross, J.B.; Speiser, D.I.; Oakley, T.H.; Patel, N.H.; Gould, D.B.; Protas, M.E. A Transcriptomic Analysis of Cave, Surface, and Hybrid Isopod Crustaceans of the Species Asellus Aquaticus. PLoS One **2015**, *10*, e0140484.

[7] Stern, D.B.; Crandall, K.A. Phototransduction Gene Expression and Evolution in Cave and Surface Crayfishes. Integr. Comp. Biol. **2018**, *58*, 398-410.

[8] Stern, D.B.; Crandall, K.A. The Evolution of Gene Expression Underlying Vision Loss in Cave Animals. Mol. Biol. Evol. **2018**, *35*, 2005-2014.

[9] Jeffery, W.R. Chapter 8. Evolution and Development in the Cavefish Astyanax. Curr. Top. Dev. Biol. **2009**, *86*, 191-221.

[10] Jeffery, W.R. Adaptive Evolution of Eye Degeneration in the Mexican Blind Cavefish. J. Hered. **2005**, *96*, 185-196.

[11] Gross, J.B.; Furterer, A.; Carlson, B.M.; Stahl, B.A. An Integrated Transcriptome-Wide Analysis of Cave and Surface Dwelling Astyanax Mexicanus. PLoS One **2013**, *8*, e55659.

[12] Hinaux, H.; Poulain, J.; Da Silva, C.; Noirot, C.; Jeffery, W.R.; Casane, D.; Retaux, S. De Novo Sequencing of Astyanax Mexicanus Surface Fish and Pachon Cavefish Transcriptomes Reveals Enrichment of Mutations in Cavefish Putative Eye Genes. PLoS One **2013**, *8*, e53553.

[13] Klaassen, H.; Wang, Y.; Adamski, K.; Rohner, N.; Kowalko, J.E. CRISPR Mutagenesis Confirms the Role of Oca2 in Melanin Pigmentation in Astyanax Mexicanus. Dev. Biol. **2018**, *441*, 313-318.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[14] Kowalko, J.E.; Ma, L.; Jeffery, W.R. Genome Editing in Astyanax Mexicanus using Transcription Activator-Like Effector Nucleases (TALENs). J. Vis. Exp. **2016**, *(112). doi*, 10.3791/54113.

[15] Ma, L.; Jeffery, W.R.; Essner, J.J.; Kowalko, J.E. Genome Editing using TALENs in Blind Mexican Cavefish, Astyanax Mexicanus. PLoS One **2015**, *10*, e0119370.

[16] Casane, D.; Retaux, S. Evolutionary Genetics of the Cavefish Astyanax Mexicanus. Adv. Genet. **2016**, *95*, 117-159.

[17] Gross, J.B.; Meyer, B.; Perkins, M. The Rise of Astyanax Cavefish. Dev. Dyn. **2015**.

[18] Protas, M.; Jeffery, W.R. Evolution and Development in Cave Animals: From Fish to Crustaceans. Wiley Interdiscip. Rev. Dev. Biol. **2012**, *1*, 823-845.

[19] Baldwin, E.; Beatty, R.A. The Pigmentation of Cavernicolous Animals: I. the Pigments of some Isopod Crustacea. Journal of Experimental Biology **1941**, *18*, 136.

[20] Verovnik, R.; Sket, B.; Trontelj, P. Phylogeography of Subterranean and Surface Populations of Water Lice Asellus Aquaticus (Crustacea: Isopoda). Mol. Ecol. **2004**, *13*, 1519-1532.

[21] Verovnik, R.; Sket, B.; Trontelj, P. The Colonization of Europe by the Freshwater Crustacean Asellus Aquaticus (Crustacea: Isopoda) Proceeded from Ancient Refugia and was Directed by Habitat Connectivity. Mol. Ecol. **2005**, *14*, 4355-4369.

[22] Verovnik, R.; Sket, B.; Prevorcnik, S.; Trontelj, P. Random Amplified Polymorphic DNA Diversity among Surface and Subterranean Populations of Asellus Aquaticus (Crustacea: Isopoda). Genetica **2003**, *119*, 155-165.

[23] Turk, S.; Sket, B.; Sarbu, ?. Comparison between some Epigean and Hypogean Populations of Asellus Aquaticus (Crustacea: Isopoda: Asellidae). Hydrobiologia **1996**, *337*, 161-170.

[24] Simona Turk-Prevorčnik; Andrej Blejec. Asellus Aquaticus Infernus, New Subspecies (Isopoda: Asellota: Asellidae), from Romanian Hypogean Waters. Journal of Crustacean Biology **1998**, *18*, 763-773.

[25] Perez-Moreno, J.L.; Balazs, G.; Wilkins, B.; Herczeg, G.; Bracken-Grissom, H.D. The Role of Isolation on Contrasting Phylogeographic Patterns in Two Cave Crustaceans. BMC Evol. Biol. **2017**, *17*, 24-9.

[26] Re, C.; Fiser, Z.; Perez, J.; Tacdol, A.; Trontelj, P.; Protas, M.E. Common Genetic Basis of Eye and Pigment Loss in Two Distinct Cave Populations of the Isopod Crustacean Asellus Aquaticus. Integr. Comp. Biol. **2018**.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[27] Protas, M.E.; Trontelj, P.; Patel, N.H. Genetic Basis of Eye and Pigment Loss in the Cave Crustacean, Asellus Aquaticus. Proc. Natl. Acad. Sci. U. S. A. **2011**, *108*, 5702-5707.

[28] Mojaddidi, H.; Fernandez, F.E.; Erickson, P.A.; Protas, M.E. Embryonic Origin and Genetic Basis of Cave Associated Phenotypes in the Isopod Crustacean Asellus Aquaticus. Sci. Rep. **2018**, *8*, 1658-8.

[29] Perez-Moreno, J.L.; Balazs, G.; Bracken-Grissom, H.D. Transcriptomic Insights into the Loss of Vision in Molnar Janos Cave's Crustaceans. Integr. Comp. Biol. **2018**, *58*, 452-464.

[30] Sket, B. Taxonomische Problematik Tier Art Asellus Aquaticus (L.) Rac. Etc. Razprave-Dissertationes SAZU, CI **1965**, *4*, 1-177.

[31] Nariai, N.; Kojima, K.; Mimori, T.; Kawai, Y.; Nagasaki, M. A Bayesian Approach for Estimating Allele-Specific Expression from RNA-Seq Data with Diploid Genomes. BMC Genomics **2016**, *17 Suppl 1*, 5.

[32] Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T. *et al.* Gene Ontology: Tool for the Unification of Biology. the Gene Ontology Consortium. Nat. Genet. **2000**, *25*, 25-29.

[33] Gene Ontology Consortium. The Gene Ontology in 2010: Extensions and Refinements. Nucleic Acids Res. **2010**, *38*, 331.

[34] Bastiani, M.J.; de Couet, H.G.; Quinn, J.M.; Karlstrom, R.O.; Kotrla, K.; Goodman, C.S.; Ball, E.E. Position-Specific Expression of the Annulin Protein during Grasshopper Embryogenesis. Dev. Biol. **1992**, *154*, 129-142.

[35] Prevorcnik, S.; Blejec A, S.B. Racial Differentiation in Asellus Aquaticus (Crustacea: Isopoda: Asellidae. Arch Hydrobiol **2004**, *160*, 193-214.

[36] Jemec, A.; Skufca, D.; Prevorcnik, S.; Fiser, Z.; Zidar, P. Comparative Study of Acetylcholinesterase and Glutathione S-Transferase Activities of Closely Related Cave and Surface Asellus Aquaticus (Isopoda: Crustacea). PLoS One **2017**, *12*, e0176746.

[37] Fiser, Z.; Prevorcnik, S.; Lozej, N.; Trontelj, P. No Need to Hide in Caves: Shelter-Seeking Behavior of Surface and Cave Ecomorphs of Asellus Aquaticus (Isopoda: Crustacea). Zoology (Jena) **2019**, *134*, 58-65.

[38] Tearle, R.G.; Belote, J.M.; McKeown, M.; Baker, B.S.; Howells, A.J. Cloning and Characterization of the Scarlet Gene of Drosophila Melanogaster. Genetics **1989**, *122*, 595-606.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[39] Bastiani, M.J.; de Couet, H.G.; Quinn, J.M.; Karlstrom, R.O.; Kotrla, K.; Goodman, C.S.; Ball, E.E. Position-Specific Expression of the Annulin Protein during Grasshopper Embryogenesis. Dev. Biol. **1992**, *154*, 129-142.

[40] Tang, L.; Tong, Y.; Cao, H.; Xie, S.; Yang, Q.; Zhang, F.; Zhu, Q.; Huang, L.; Lu, Q.; Yang, Y. *et al.* The MTMR9 rs2293855 Polymorphism is Associated with Glucose Tolerance, Insulin Secretion, Insulin Sensitivity and Increased Risk of Prediabetes. Gene **2014**, *546*, 150-155.

[41] Yanagiya, T.; Tanabe, A.; Iida, A.; Saito, S.; Sekine, A.; Takahashi, A.; Tsunoda, T.; Kamohara, S.; Nakata, Y.; Kotani, K. *et al.* Association of Single-Nucleotide Polymorphisms in MTMR9 Gene with Obesity. Hum. Mol. Genet. **2007**, *16*, 3017-3026.

[42] Aspiras, A.C.; Rohner, N.; Martineau, B.; Borowsky, R.L.; Tabin, C.J. Melanocortin 4 Receptor Mutations Contribute to the Adaptation of Cavefish to Nutrient-Poor Conditions. Proc. Natl. Acad. Sci. U. S. A. **2015**, *112*, 9668-9673.

[43] Riddle, M.R.; Aspiras, A.C.; Gaudenz, K.; Peuss, R.; Sung, J.Y.; Martineau, B.; Peavey, M.; Box, A.C.; Tabin, J.A.; McGaugh, S. *et al.* Insulin Resistance in Cavefish as an Adaptation to a Nutrient-Limited Environment. Nature **2018**, *555*, 647-651.

[44] Rajput, R.; Ahlawat, P. Prevalence and Predictors of Non-Alcoholic Fatty Liver Disease in Prediabetes. Diabetes Metab. Syndr. **2019**, *13*, 2957-2960.

[45] Berning, D.; Adams, H.; Luc, H.; Gross, J.B. In-Frame Indel Mutations in the Genome of the Blind Mexican Cavefish, Astyanax Mexicanus. Genome Biol. Evol. **2019**, *11*, 2563-2573.

[46] Gross, J.B.; Borowsky, R.; Tabin, C.J. A Novel Role for Mc1r in the Parallel Evolution of Depigmentation in Independent Populations of the Cavefish Astyanax Mexicanus. PLoS Genet. **2009**, *5*, e1000326.

[47] Gross, J.B.; Wilkens, H. Albinism in Phylogenetically and Geographically Distinct Populations of Astyanax Cavefish Arises through the Same Loss-of-Function Oca2 Allele. Heredity (Edinb) **2013**, *111*, 122-130.

[48] Riddle, M.R.; Aspiras, A.C.; Gaudenz, K.; Peuss, R.; Sung, J.Y.; Martineau, B.; Peavey, M.; Box, A.C.; Tabin, J.A.; McGaugh, S. *et al.* Insulin Resistance in Cavefish as an Adaptation to a Nutrient-Limited Environment. Nature **2018**, *555*, 647-651.

[49] Protas, M.E.; Hersey, C.; Kochanek, D.; Zhou, Y.; Wilkens, H.; Jeffery, W.R.; Zon, L.I.; Borowsky, R.; Tabin, C.J. Genetic Analysis of Cavefish Reveals Molecular Convergence in the Evolution of Albinism. Nat. Genet. **2006**, *38*, 107-111.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[50] Elipot, Y.; Hinaux, H.; Callebert, J.; Launay, J.M.; Blin, M.; Retaux, S. A Mutation in the Enzyme Monoamine Oxidase Explains Part of the Astyanax Cavefish Behavioural Syndrome. Nat. Commun. **2014**, *5*, 3647.

[51] Stahl, B.A.; Gross, J.B. Alterations in Mc1r Gene Expression are Associated with Regressive Pigmentation in Astyanax Cavefish. Dev. Genes Evol. **2015**, *225*, 367-375.

[52] Wray, G.A. The Evolutionary Significance of Cis-Regulatory Mutations. Nat. Rev. Genet. **2007**, *8*, 206-216.

[53] Signor, S.A.; Nuzhdin, S.V. The Evolution of Gene Expression in Cis and Trans. Trends Genet. **2018**, *34*, 532-544.

[54] Buchberger, E.; Reis, M.; Lu, T.H.; Posnien, N. Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies. Genes (Basel) **2019**, *10*, 10.3390/genes10070492.

[55] Fraser, H.B. Improving Estimates of Compensatory Cis-Trans Regulatory Divergence. Trends Genet. **2019**, *35*, 3-5.

[56] Lawson, H.A.; Cheverud, J.M.; Wolf, J.B. Genomic Imprinting and Parent-of-Origin Effects on Complex Traits. Nat. Rev. Genet. **2013**, *14*, 609-617.

[57] Speiser, D.I.; Pankey, M.S.; Zaharoff, A.K.; Battelle, B.A.; Bracken-Grissom, H.D.; Breinholt, J.W.; Bybee, S.M.; Cronin, T.W.; Garm, A.; Lindgren, A.R. *et al.* Using Phylogenetically-Informed Annotation (PIA) to Search for Light-Interacting Genes in Transcriptomes from Non-Model Organisms. BMC Bioinformatics **2014**, *15*, 35-x.

[58] Stahl, B.A.; Gross, J.B.; Speiser, D.I.; Oakley, T.H.; Patel, N.H.; Gould, D.B.; Protas, M.E. A Transcriptomic Analysis of Cave, Surface, and Hybrid Isopod Crustaceans of the Species Asellus Aquaticus. PLoS One **2015**, *10*, e0140484.

[59] Carlini, D.B.; Fong, D.W. The Transcriptomes of Cave and Surface Populations of Gammarus Minus (Crustacea: Amphipoda) Provide Evidence for Positive Selection on Cave Downregulated Transcripts. PLoS One **2017**, *12*, e0186173.

[60] Stern, D.B.; Crandall, K.A. The Evolution of Gene Expression Underlying Vision Loss in Cave Animals. Mol. Biol. Evol. **2018**, *35*, 2005-2014.

[61] Stern, D.B.; Crandall, K.A. Phototransduction Gene Expression and Evolution in Cave and Surface Crayfishes. Integr. Comp. Biol. **2018**, *58*, 398-410.

[62] Huang, Z.; Titus, T.; Postlethwait, J.H.; Meng, F. Eye Degeneration and Loss of Otx5b Expression in the Cavefish Sinocyclocheilus Tileihornes. J. Mol. Evol. **2019**, *87*, 199-208.

bioRxiv preprint doi: https://doi.org/10.1101/845990; this version posted November 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

781 [63] Friedrich, M.; Chen, R.; Daines, B.; Bao, R.; Caravas, J.; Rai, P.K.; Zagmajster, M.;
782 Peck, S.B. Phototransduction and Clock Gene Expression in the Troglobiont Beetle
783 Ptomaphagus Hirtus of Mammoth Cave. J. Exp. Biol. **2011**, *214*, 3532-3541.

784 [64] Tierney, S.M.; Langille, B.; Humphreys, W.F.; Austin, A.D.; Cooper, S.J.B. Massive
785 Parallel Regression: A Precis of Genetic Mechanisms for Vision Loss in Diving Beetles.
786 Integr. Comp. Biol. **2018**, *58*, 465-479.

787 [65] Lefebure, T.; Morvan, C.; Malard, F.; Francois, C.; Konecny-Dupre, L.; Gueguen, L.;
788 Weiss-Gayet, M.; Seguin-Orlando, A.; Ermini, L.; Sarkissian, C. *et al*. Less Effective
789 Selection Leads to Larger Genomes. Genome Res. **2017**, *27*, 1016-1028.

790 [66] Konec, M.; Prevorcnik, S.; Sarbu, S.M.; Verovnik, R.; Trontelj, P. Parallels between
791 Two Geographically and Ecologically Disparate Cave Invasions by the Same Species,
792 Asellus Aquaticus (Isopoda, Crustacea). J. Evol. Biol. **2015**, *28*, 864-875.

793