**3D genome organization under extreme metabolic adaptation shows variations in genome compartmentalization and looping**

Tathagata Biswas[1], Hua Li[1], Nicolas Rohner[1,2,*]

**Affiliation**

[1] Stowers Institute for Medical Research, Kansas City, MO 64110, USA

[2] Department of Cell Biology & Physiology, University of Kansas Medical Center, Kansas City, KS 66103, USA

*Email for correspondence: nro@stowers.org

**Abstract**

Three-dimensional (3D) genome organization plays a crucial role in gene regulation by influencing cellular functions, such as bringing regulatory elements like enhancers into proximity with their target genes. In this study, we explore variations in 3D genome organization within different morphs of the fish *Astyanax mexicanus*, which includes a river-dwelling surface fish and multiple cave-dwelling morphs. The cave morphs, also known as cavefish, have adapted to the dark and nutrient-scarce cave environments, resulting in distinct metabolic reprogramming that has made them persistently hyperglycemic, resilient to starvation, and prone to fat accumulation. Focusing on metabolic evolution in cavefish, we used liver tissue to conduct genome-wide contact mapping through Hi-C experiments. Comparing the 3D genome architecture of two cave morphs – Pachón and Tinaja – with that of surface fish, we observed significant changes in genome compartmentalization and genomic loops. Some of these changes were common in both Pachón and Tinaja and were classified as cave-specific looping and compartmental switches. By integrating liver Hi-C data with RNA-seq, ATAC-seq, and histone ChIP-seq datasets, we gained deeper insight into the relationship between cave-specific 3D genome features and the transcriptional activity at associated loci. Additionally, we identified structural variations in the form of common genomic inversions in Pachón and Tinaja compared to surface fish. Therefore, this study is the first to define the 3D genome organization of *Astyanax mexicanus* in livers and highlights how adaptation to extreme environments is associated with significant variations in 3D genome architecture, even within the same species.

## Introduction

Three-dimensional (3D) packaging of the genome inside the nucleus has significant influence over several essential cellular and tissue functions. 3D genome organization affects physical proximity of different regulatory elements to their respective genes, which is essential for effective gene regulation [1-3]. In the 3D space of the nucleus, genome looping, at a scale ranging from a few to some hundreds of kilobases, ensures close contact between long-distance cis-regulatory elements, like enhancers, and their corresponding effector gene loci for effective gene regulation. The 3D looping of a genome is primarily mediated through the loop extrusion processes by cohesion/CTCF associated structural maintenance of chromosomes (SMC) protein complexes [4-6]. The 3D looping also allows genes separated over large genomic distances in the linear genome to share common enhancers, thereby aiding co-regulation of genes within the topologically associating domains (TADS) created by one or many such loops [7, 8]. With multitude of loops and interactions occurring all over the genome, when observed on larger genomic scale, the genomic DNA can also be broadly divided into active (A) and inactive (B) compartments, where loci with relatively higher molecular activity and gene density interact more with each other and segregate together in the 3D space to form the 'A' compartment, compared to the other compartment with lesser molecular activity and is termed as 'B'. While alterations in the genome architecture affect molecular machinery regulating transcriptomic profiles and diseases [9-11], some of the recent studies have shown that gene expression is mostly resilient to perturbations in the 3D genome organization [12, 13] and suggested that 3D genome structures aided precise gene regulation, but were not entirely essential [12]. One possible way to resolve this paradox would be to investigate the significance of 3D genome organization across more organisms and conditions. In this regard, researchers have conducted a wide range of evolutionary studies that show 3D genome organization is highly conserved across

syntenic regions in different species [14-16]. However, while these conserved syntenic regions are well studied, analyzing 3D organizational variations among different species is challenging due to inherent inter-species noise. To get a deep insight into 3D genome changes and variations, it is essential to compare within the same species. Therefore, in this study we take advantage of the fish species *Astyanax mexicanus* – which, while still being the same species, exists naturally as two different morphs – to investigate 3D genome dynamics in between morphs of the same species evolving under two extremely different environmental conditions.

Approximately 200,000 years ago, the ancestral river-dwelling 'surface' *Astyanax* colonized multiple caves in Mexico through multiple independent events of invasion, giving rise to multiple cave morphs like Pachón and Tinaja [17, 18]. The caves, with their impenetrable darkness, scarcity of nutrients, are a great example of extreme environment. The cave morphs of *Astyanax*, collectively termed as cavefish, evolved conspicuous cave-adaptive traits, like loss of eye, pigmentation and resilience to a wide-range of metabolic stresses [19-25]. The cavefish show remarkable changes in their metabolism with respect to surface fish and are resilient to starvation, while being persistently hyperglycemic, hyperphagic, and exhibit enhanced lipogenesis. Therefore, in this study, we focused on the metabolically important tissue – the liver – to investigate the 3D genome architecture between surface and cavefish (Fig. 1a). We performed genome-wide contact analysis through Hi-C experiment and established the first 3D genome organizational map of *Astyanax mexicanus* using surface fish and two different cave morphs – Pachón and Tinaja. Despite being the same species, we could identify significant variations in the 3D genome organization in between surface and cavefish morphs in terms of their genome compartments, genomic loops, and genomic inversions. Comparing the two cave morphs with surface fish, we could also identify, for the first time, the common changes in the 3D genome features associated with evolving under

extreme metabolic stress. Finally, these common variations in 3D genome organization included loci containing genes involved in pathways critical for the metabolic adaptation of cavefish. Subsequently, these genes are also the candidate genes, potentially regulating metabolic resilience in general, paving way for future therapeutic studies.

## Materials and Methods

### Sample collection for Hi-C

Liver tissues were collected from three adult fish (approximately 500 days old) from each morph: Pachón, Tinaja, and surface fish. After a quick 1xPBS wash, whole livers were flash frozen in liquid nitrogen. The frozen tissues were homogenized with help of Zirconium beads (3mm) in a Beadbug 6 microtube homogenizer at 4000 osc/min for 30secs. Thereafter, Hi-C data was generated using the Arima-HiC kit (PN: A510008). After necessary QC, sample where ready for library preparation.

### Libraries and Sequencing

For Pachón and surface fish: HiC libraries were generated according to manufacturer's directions for the Arima-HiC Kit: Library Preparation using KAPA Hyper Prep Kit protocol (Document Part No. A160139 v00) starting with 1.95-5.59ug of large proximally-ligated DNA molecules for Pachón and surface, and with 0.56-0.81 ug of large proximally-ligated DNA molecules for Tinaja, as assessed using the Qubit Fluorometer (Life Technologies). Libraries were prepared using the S220 Focused-ultrasonicator (Covaris) to shear samples to 400bp, and the KAPA Hyper Prep Kit (Roche Cat #: KK8500/7962312001) with Illumina IDT-TruSeq DNA UD Indexes (Illumina Cat #: 20020590) following Arima's modified protocol with 10 cycles of library PCR amplification.

Resulting short fragment libraries were checked for quality and quantity using the Bioanalyzer (Agilent) and Qubit Fluorometer (Life Technologies). For Pachón and surface, libraries were pooled, re-quantified and sequenced as 150bp paired reads on an S1 flow cell (v1.5) on the Illumina NovaSeq 6000 instrument utilizing RTA and instrument software versions current at the time of processing. For Tinaja, libraries were pooled, re-quantified and sequenced as 150bp paired reads on a P3 flow cell on the Illumina Nextseq 2000 instrument utilizing RTA and instrument software versions current at the time of processing.

**HiC-seq data process**

Briefly, reads were processed using Juicer from Aiden lab with the default parameters [26], against the reference genome Amex_1.1. Final contact databases were created using mapq >= 30.

A total of 620 million to 830 million reads were obtained per sample. Following quality filtering, the number of valid read pairs ranged from 189 million to 419 million.

**A/B compartment analysis**

A/B compartment analysis was performed at 25kb resolution using POSSUMM [27]. The choice of the first or the second PCA depended on how it corresponded to the plaid-like pattern observed in the Hi-C data. Once the PCA is decided, the signs of these values are determined to make sure that the positive values always correspond to the active state, which is represented by the high gene density. The significant A/B bin switch are identified using lm() function in base R package. We intersect the gene model with A/B bins using bedtools intersect.

**Loop analysis using SIP and CTCF motif**

Individual .hic files within each morph were merged to form one .hic file for each of the three *Astyanax* morphs. Loops were identified using the default parameters with 5k, 10k, 15k, and 20k resolutions. Results were then combined to generate a final loops file for each sample. We have identified 2571 loops in surface, 3510 loops in Pachon, 1732 loops in Tinaja. Loops thar are wider than 1MB are excluded. MotifFinder was used to find CTCF motifs in the anchors points of the loops. MotifFinder: https://github.com/aidenlab/juicer/wiki/MotifFinder.

**Aligning other sequencing data with Hi-C**

ATAC-seq, Chip-seq and RNA-seq data were downloaded from GSE153052. ATAC-seq data was aligned to AMEX_1.1 using bowtie2. MACS2 and then IDR was used to call peaks for each replicate. Reads were aligned to AMEX_1.1 genome and the gene model was mapped from astMex2 Ensembl106 gene model, using liftoff [28] and STAR. The gene expression values TPM were estimated using RSEM. TPMs were normalized using the quantile normalization method to remove any potential batch effect. Differential gene expression analysis was done using R package limma [29].

**Results**

**Hi-C based 3D genome organization in surface fish, Pachón, Tinaja**

To determine the 3D genome architecture of *Astyanax* livers, we isolated livers from adult Pachón, Tinaja, and surface fish for Hi-C based analysis. Hi-C data from each of the samples were generated using the Arima Hi-C kit, resulting in restriction enzyme cuts at GATC and GANTC motifs across the genome. With high quality sequencing of the library and then merging of the biological replicates for each of the three morphs of *Astyanax*, we obtained 1348 million valid

reads for Hi-C in surface, 1093 million valid reads in Pachón and 603 million valid reads in Tinaja with an average of 304 million contacts across each morph. With this data, our resolution for the contact matrix could go down to 5000 bp. Each of the genome level Hi-C map generated after aligning with AMEX_1.1 genome clearly reflected the presence of the 25 chromosomes, with high intra-chromosomal interactions (Fig. 1b). While the biological replicates within each morph showed high degree of correlation, the three different morphs showed an overall good correlation, which is expected, since they are still the same species.

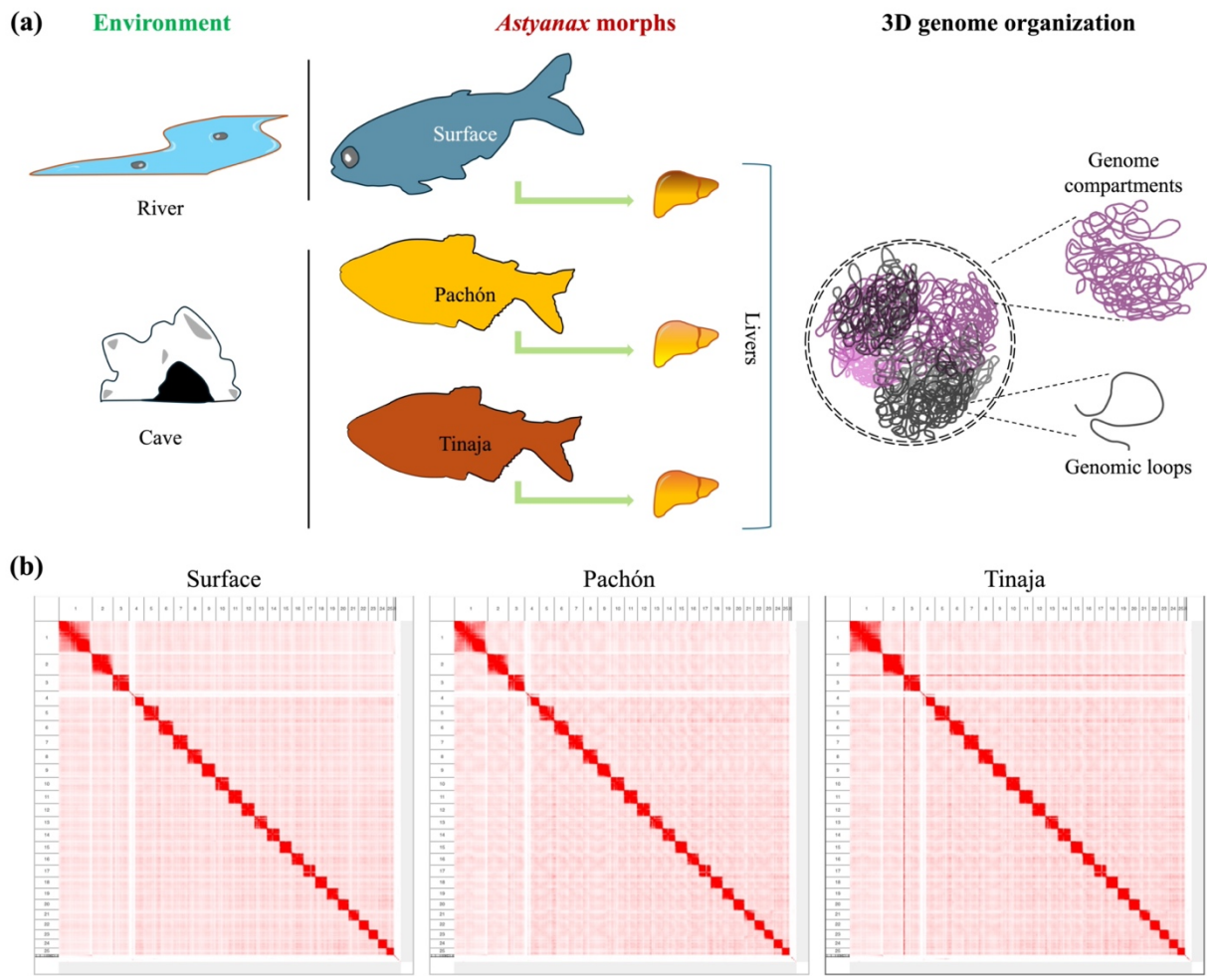



Fig.1: (a) Hi-C experiment was performed on the livers of surface and cave morphs of *Astyanax mexicanus* to investigate the changes in 3D genome organization within the same species but living under different environmental conditions. (b) Whole genome 3D contact matrix derived from liver Hi-C experiment on surface, Pachón and Tinaja fish morphs.

**3D genome segregation into active (A) and inactive (B) compartments in *Astyanax***

Hi-C analysis in different species have shown that the genome inside the nucleus segregates broadly into two types of 3D compartments. One compartment, characterized by higher gene density and transcriptomic activity, leads to a higher frequency of contact between different loci, forming the active 'A' compartment and segregates separately from the relatively inactive 'B' compartment [30]. This compartmentalization of the genome is reflected in the plaid-like pattern of the Hi-C contact matrix, which when put under principal component analysis, the principal eigenvector and eigenvector sign changes define the A and B compartmental boundaries. In this study, observing the Hi-C contact map, especially at the chromosomal scale, of each of the morphs of *Astyanax* showed clear emergence of plaid-like pattern, reminiscent of A/B compartmentalization of the genome (Fig. 2a, Supplementary Fig. 1). To define the compartments, we employed Principal component analysis (PCA) Of Sparse, SUper Massive Matrices (POSSUMM) [27]. Using POSSUMM, in all the three morphs, we defined the active/inactive compartments in *Astyanax* (magenta track depicting the eigenvectors in Fig. 2a) with the positive values corresponding to the active state with high gene density. To check if indeed the segregation of the genome into A and B compartments reflected differential transcriptional activity, we compared the transcript reads of coding genes of the respective compartments by aligning the compartment data with data from liver RNA-seq [31]. In line with the compartmental segregation, overall, across all the three morphs, the A compartment showed higher transcriptional activity than the B compartment (Fig. 2b).

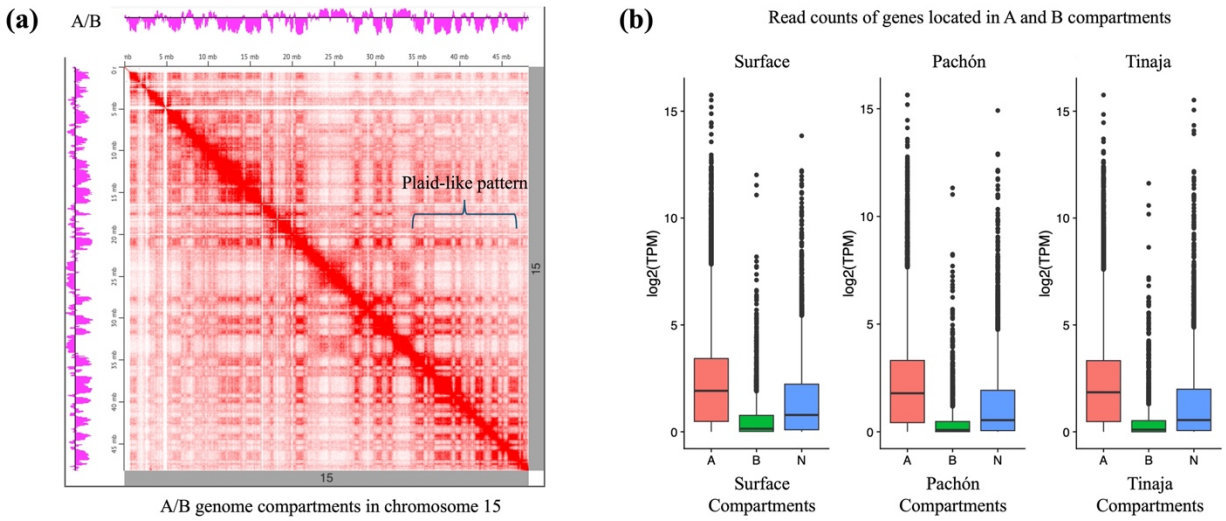



Fig. 2: (a) Depiction of the plaid-like pattern in the chromosome level contact frequency map of chromosome 15 in Pachón fish. A/B compartment, in-sync with the plaid-like pattern, is the magenta track adjacent to the axes of the matrix. The positive peaks being A and negatives being B compartment. (b) Active 'A' compartment in the genome shows higher transcriptional activity across all the three morphs of *Astyanax* as compared to the inactive 'B' compartment. N being the neutral compartments.

**Switching of the genome compartments in cavefish**

Adapting to an extreme environment is associated with large scale transcriptomic and regulatory changes in the cavefish [31, 32]. Since active A and inactive B compartments associated well with transcriptional state, we wanted to investigate if adaptation to the caves associated with changed compartmentalization of the genome in cave morphs with respect to the surface. Comparative analysis of the A and B compartments, in between the two cave morphs with surface, revealed that while being mostly similar, multiple regions in the genome underwent compartmental switching in between A and B. Out of 53988 total bins of 25k, 665 bins exhibited A/B switching (pval<=0.001). As anticipated, switched genome compartments also mirrored altered transcriptomic trend among the genes located within the switched loci in both Pachón and Tinaja. Genes moving to A in cavefish showed higher median fold change, as compared to the ones switching to B (Fig. 3a, Supplementary Fig. 2) when compared to surface, while the genes in the

regions that remained in the same compartment across cave morphs and surface, showed no such change in transcriptional trend. Thus, cave adaptation in *Astyanax* associated with re-organization of active - inactive compartments, which simultaneously echoed their overall transcriptional states.

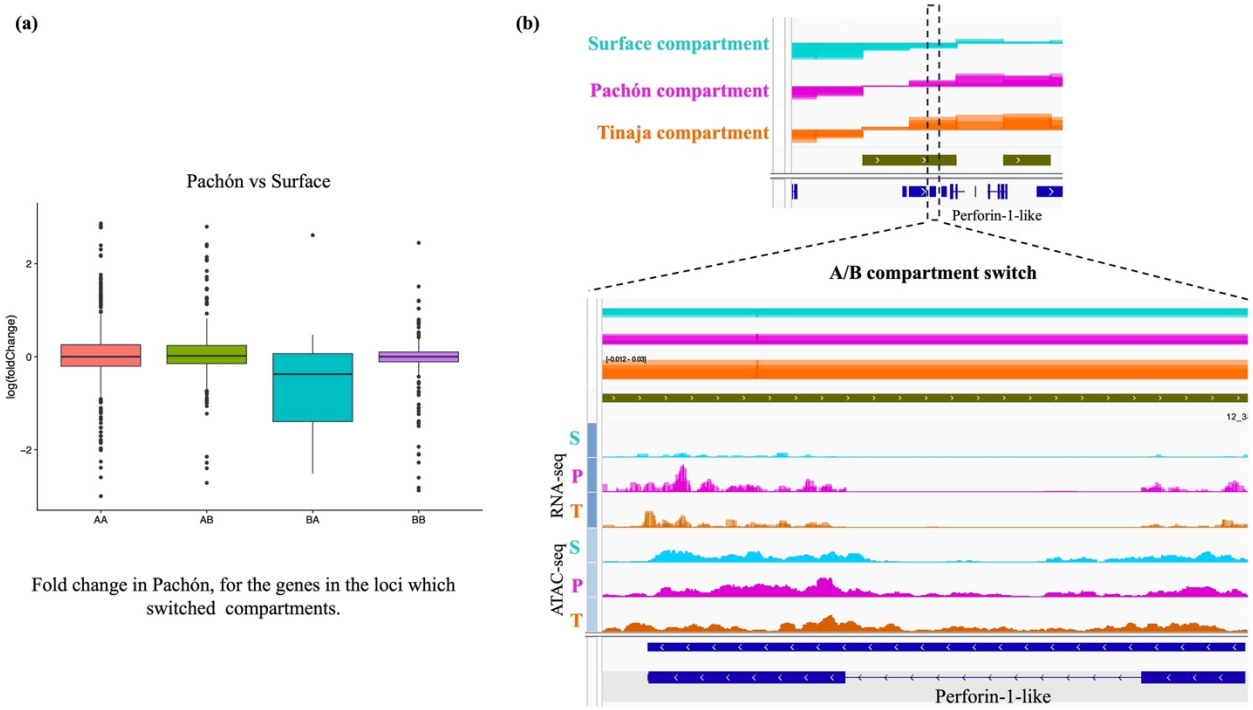



Fig. 3: (a) Boxplot depicting fold change in Pachón, for the genes in the loci which switched compartments. AA and BB denotes unchanged compartments, AB denotes loci which is active in Pachón and inactive in surface, BA denotes loci active in surface and inactive in Pachón. (b) Example of a BAA loci, where B compartment loci, in surface fish, switches to active A compartment in Pachón and Tinaja cave morphs. The switched loci comprise of genes, including Perforin-1-like (dashed inset), which when aligned to the RNA-seq profile, shows higher levels of expression in both the cavefish as compared to surface.

**Emergence of cave-specific compartmentalization in *Astyanax mexicanus***

Among the range of compartmental switches, the ones which are critically associated with cave adaptation should be evident in both Pachón and Tinaja. Therefore, we next identified all the compartmental switches which were present in both cave morphs. Notably, these are the first instances of common variations of 3D genome organization associated with extreme cave adaptation. The switch from B compartment, in surface fish, to the A compartment in both Pachón

and Tinaja cavefish (termed as BAA regions) comprised of 287 genes under the loci (Fig. 3b). Similarly, 36 genes were located in the loci who switched from the A compartment in surface fish to the inactive B compartment in both the cave morphs (termed as ABB, regions). To explore if the conserved variations in cave morphs associated with biological processes important for cave adaptation, we performed Gene Ontology (GO) term analysis (Supplementary Table 1). For the BAA regions, we observed enrichment for metabolic pathways like "chondroitin sulfate proteoglycan metabolic process", "pyrimidine nucleoside monophosphate metabolic process", "pyrimidine nucleoside monophosphate biosynthetic process" and more. The fact that the BAA regions include genes which contribute to pyrimidine related pathways is significant because, as previously reported, another intermediate in pyrimidine synthesis, orotic acid, is already known to be sharply responsive to starvation and refeeding, specifically in the *Astyanax* liver tissue [20]. Additionally, we also observed the presence of non-metabolic terms, associated with eye and retinal development like "retina development in camera-type eye" and "compound eye morphogenesis". This strongly suggests a more systemic change reflected in the genomic compartmental switches of the blind cavefish, one that is clearly not restricted to the liver or metabolism.

**Genomic contact loops in surface and cave morphs of *Astyanax***

While the A/B compartmental analysis occurs at a more global scale and explains overall active and inactive regions of the genome, to understand the transcriptional regulation of the genes in context of 3D genome contacts, it is important to understand the interaction between loci at a smaller scale and higher resolution. Analyzing the genomic contacts identifies 3D interactions in the nuclear space between coding regions and regulatory regions, each lying at a range of different

distances in an otherwise linear genome. Therefore, we investigated our Hi-C data for the presence of significant genomic loops in the surface and the two cavefish morphs. The *Astyanax* being a non-mammalian system, we used the Significant Interaction Peak caller (SIP) method to detect relevant genomic loops as it is equally effective in detecting loops in humans and other organism, including invertebrates [33].

We analyzed the loops under three different resolution settings (5k, 10k, 15k, and 20k) in all the three morphs to identify a total of 2525 loops in surface fish, 3459 loops in Pachón and 1690 loops in Tinaja from their respective liver Hi-C data. While the loops were determined by the SIP pipeline, we also overlapped the anchor points of the identified loops with binding motifs of CTCF, a known regulator of genomic loop extrusion [6]. The alignment analysis revealed that majority of the loop anchor points which coincided with CTCF binding motifs, were of convergent CTCF motifs (Fig. 4a). While there were also a few loops which had tandemly placed CTCF motif at their anchor points, but divergent CTCF motifs on anchor points were very rare (Fig. 4a). Comparing the loops in surface, Pachón and Tinaja we observed that they shared mostly similar looping loci and architecture for a major part of their 3D genome organization, with over 1500 loops identified as being commonly present in all the three morphs. However, we were also able to identify loops which were found only in either cavefish or surface fish. We identified a total of 231 surface-specific loops, which were absent in both the cave populations, while both Pachón and Tinaja had common cave-specific loops numbering to around 200.

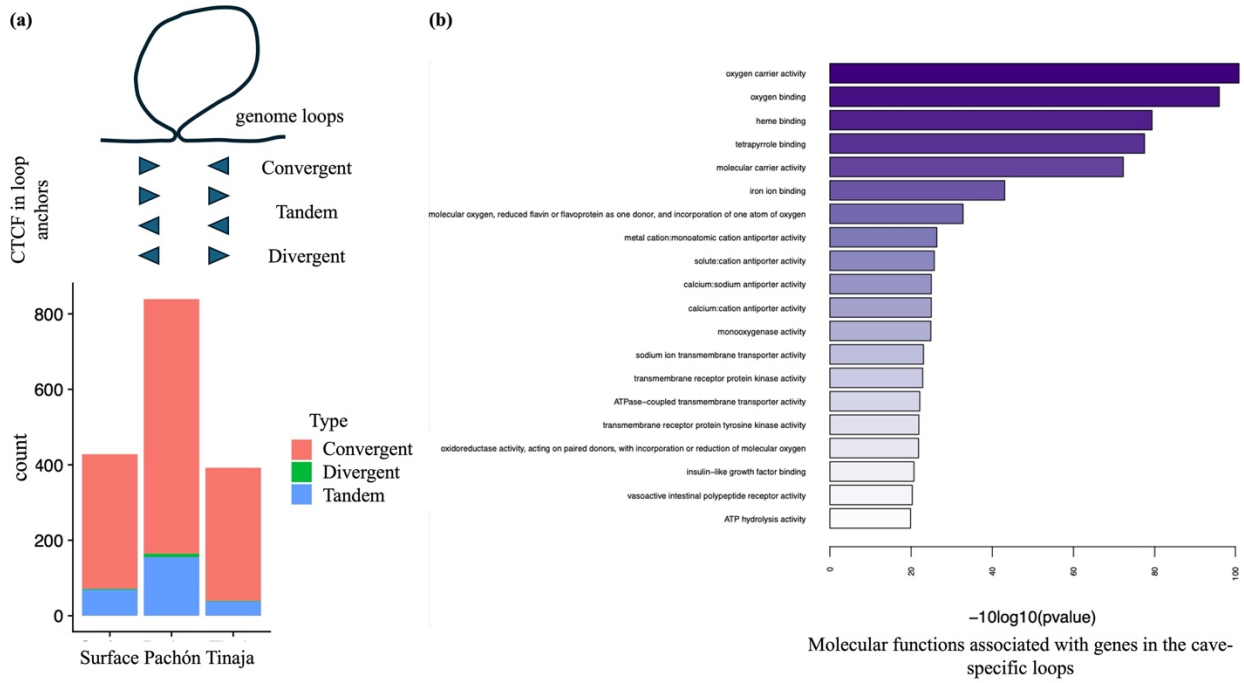



Fig. 4: (a) Distribution of CTCF anchor points for loops in each of the three morphs. The CTCF sites were either convergent, divergent or in tandem, as depicted in the cartoon. (b) Gene Ontology analysis show that the molecular functions associated with the genes of cave-specific loop are associated with cavefish-metabolism.

**Genes and pathways associated with the cave-specific loops**

Listing the genes located under the morph specific loop intervals, we identified a total of 667 genes associated with the surface-specific loops and 620 genes were under the cave-specific loop intervals and loop anchor points. An overall GO term analysis showed that many of the molecular functions associated with the genes of cave-specific loop were specifically associated with known aspects of cavefish-metabolism (Supplementary Table 2, Fig. 4b). The loop interval genes enriched for terms associated with oxidative activities, like “oxygen carrier activity”, “oxygen binding”, “oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen” and more. Previous studies have also shown that cavefish and surface fish show differences in their oxidative metabolism [20], which is also observed in *in vitro* set up [34]. Another important pathway enriched was that of “heme binding” which included the Hemoglobin subunit

beta-1-like gene. Notably, cavefish have been previously shown to have higher expression of hemoglobin related genes early during development [35]. Finally, the genes also enriched for "insulin−like growth factor binding" which included the genes *ccn2a* and *ccn1l2*.

**Altered transcriptomic profile under morph-specific loops**

Changes in the genome looping is strongly associated with changes in the transcriptomic profiles. Therefore, we went further and integrated gene expression data, chromatin accessibility and epigenetic signatures with our Hi-C data. We aligned surface fish, Pachón and Tinaja liver RNA-seq, ATAC-seq and histone ChIP-seq datasets [31] with our Hi-C data, to assess the possible role of changed genome looping on cave adaptation. From the list of 620 genes under cave-specific loops, 117 (p-value <=0.01) show gene expression among the three morphs. Of these 117 which show expression, 75 genes show higher expression in cavefish as compared to surface fish (Supplementary Table 3). A similar analysis of the surface-specific loops rendered a total of 146 genes (p-value <=0.01) out of the total 667, which were expressed among the three morphs and only 102 of them showed higher expression in surface fish compared to the cave morphs (Supplementary Table 4).

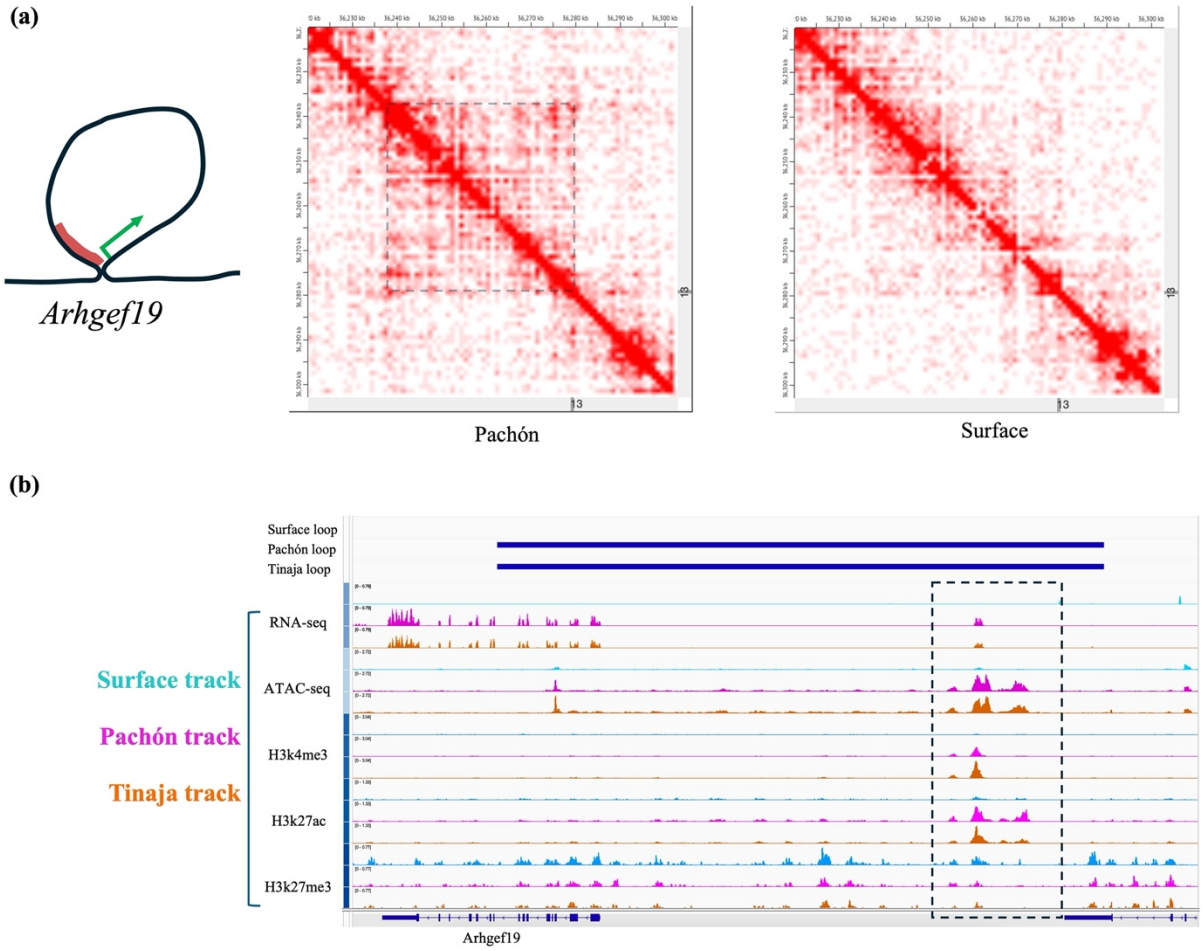



Fig. 5: (a) Cave-specific loop of *Arhgef19*, here depicted in Pachón and surface in chromosome 13. Dashed box marks the loop in the Pachón contact matrix. (b) Aligning the RNA-seq, ATAC-seq and signature histone ChIP-seq data with the loop calling dataset to observe activity around the loop anchor points. The box marks the putative bivalent regulatory region near the other anchor point of the 40kb long loop, away from *Arhgef19*.

Within the loops, the loop anchor points, commonly enriched with enhancers and promoters, are the most likely regions to affect transcriptomic profile [36, 37]. Therefore, among the 75 genes which were under the cave-specific loop intervals and upregulated in cavefish, we focused on the genes which were located near loop anchor points. We found two genes and associated loops where the genes showed significant differential expression and were also located near the anchor points – *Arhgef19* (Rho guanine nucleotide exchange factor 19) and *Endog* (endonuclease G). While *Arhgef19* was associated with a 40kb long (13:36,235,001-36,275,000) cave-specific loop on

chromosome 13 (Fig. 5a), the gene *Endog* was located near the anchor point of an approximately 145kb long (12:1,555,001-1,700,000) cave-specific loop on chromosome 12 (Supplementary Fig. 3). Additionally, *Arfgef19* also showed the highest fold change in expression among the 75 shortlisted genes when both Pachón and Tinaja were compared to surface fish. Interestingly, while *Arfgef19* was located at one end of the cave-specific loop, whereas the other end of the loop included spans a genomic region which when aligned with histone Chip-seq data and ATAC-seq data showed signatures of being a cis-regulatory loci. The genomic spans with H3K4me3 (trimethylation of histone H3 at lysine 4) mark are considered promoters, while the active enhancers are the ones with H3K27ac (acetylation of histone H3 at lysine 27) marks. Interestingly, the regions with both H3K4me3 and H3K27me3 (trimethylation of histone H3 at lysine 27) marks are potentially the bivalent domains (dashed line box, Fig. 5b) [38, 39]. Therefore, with certain regions of the other end of the loop showing highly accessible regions with strong acetylation marks of H3K27ac in cavefish while being absent in surface, the cave-specific loop associated with *Arhgef19* potentially plays a critical regulatory role. Further, the roles of both *Arhgef19* and *Endog*, in metabolic response to stress, and possibly resilience, could also be conserved across species. A recent collaborative study involving our group showed that inhibition of a fatty acid transporter *slc27a2a* rescued detrimental effect of steatosis and atrophy of the 48h starved zebrafish embryo livers [21]. Intriguingly, the list of genes differentially expressed between the simply starved zebrafish and those starved but later rescued included both *Arhgef19* and *Endog*. Further, previous studies also show role of *Arhgef19* and *Endog* in metabolism, particularly their association with adipocyte differentiation and response to high-fat diet [40-42].

**Hi-C contact maps reveal inversions in the cavefish genome**

Cavefish genomes exhibit structural variations with respect to surface fish and several instances of insertions and deletions were identified across multiple cave morphs, with average cumulative size of cave morph deletions and insertions being 19.4 Mb [43]. In addition to insertions and deletions, inversions are a major type of structural variations which are known to affect recombination and are linked to developing fitness in different ecotypes of a species evolving under different local environments [44, 45]. Analyzing the 3D genome Hi-C data has the potential to identify such structural variations of inversions in the genome. Using the deep-learning based framework of EagleC pipeline we analyzed our liver Hi-C data to identify 37 genomic inversions across surface and cave morphs. One such example of inversion common across the two cave morphs with respect to surface was found on chromosome 1, spanning across a length of 2.7Mb and having the clear "butterfly" pattern, which is associated with genomic inversions (Fig. 6). The genes in the inverted region (Supplementary Table 5) do not necessarily belong to pathways associated directly with metabolism, apart from the gene *gltpd2*, which is associated with lipid metabolism. Instead, a large fraction of the gene in this inversion were known regulators of cell cycle. Considering cell cycle regulation is an essential aspect of resilience to physiological stressors and countering tissue atrophy [21], it is interesting to note the presence of these cell cycle genes inside the inverted locus.

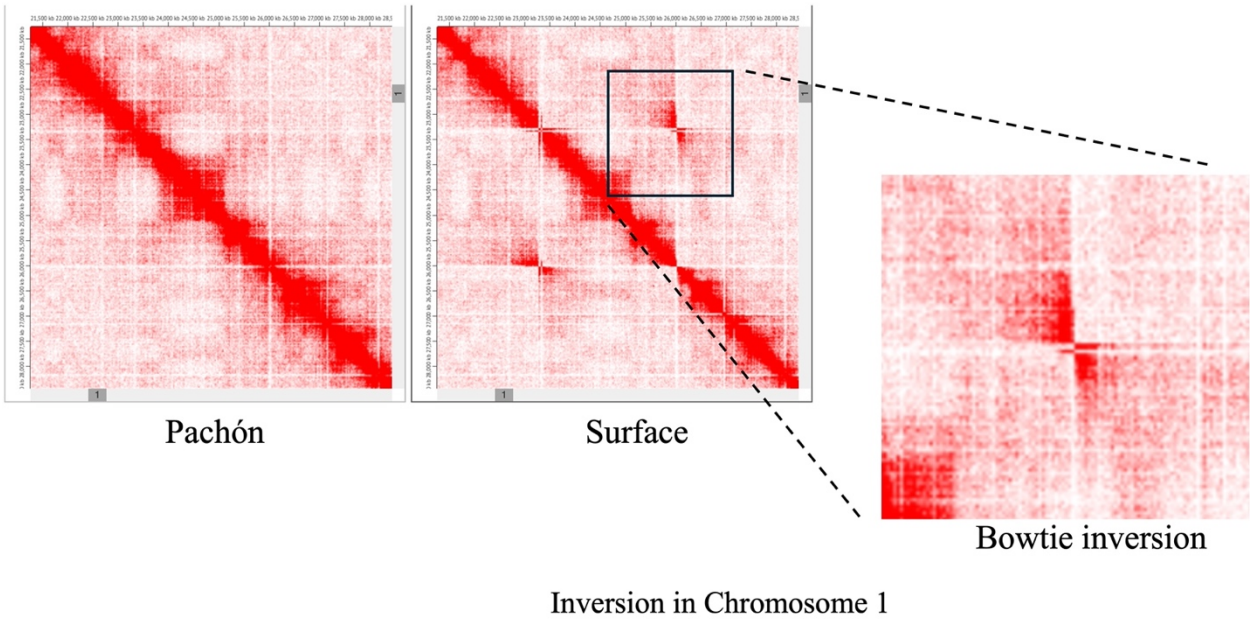



Fig. 6: A genomic inversion in between morphs of *Astyanax* in chromosome 1, found in both Pachón and Tinaja with respect to the surface fish (here depicted only comparing Pachón and surface fish). Inset highlights the signature "butterfly" pattern in the Hi-C contact map signifying genomic inversion.

**Discussion**

In this study, we investigated the 3D genome architecture in the livers of surface and cave morphs of *Astyanax mexicanus*, establishing the organization of the genome into compartments and loops. We found that significant changes occur in both compartmental organization and looping architecture when comparing cave and surface morphs, despite them being the same species. Additionally, the loci of common 3D genome variations in the livers of the two cave morphs, relative to the surface fish, include genes involved in pathways related to cave adaptation. Notably, most of these genes and pathways associated with cave-specific 3D variations are metabolic in nature, suggesting that the 3D genome architecture in *Astyanax* is largely tissue-specific.

While comparing 3D looping variations with gene expression data helped us identify candidate genes, such as *Arhgef19* and *Endog*, associated with the unique metabolism of cavefish, it is

interesting to note that a substantial number of genes within regions of 3D looping variation do not show changes in transcriptional activity (only 75 genes showed higher expression in cavefish, out of 620 genes located in cave-specific loops). This supports the idea that not all 3D looping changes result in an immediate effect on gene transcription. It is likely that the altered architecture may poise and prime many of these genes for a more dynamic response when exposed to extreme stimuli, such as metabolic stress [46]. Further investigation, using *in vivo* and *in vitro* [34, 47] systems, under metabolic stimuli, would help us understand the role of these candidate genes in metabolic stress response.

Another advantage of a detailed Hi-C interaction map is its potential to identify regulatory regions for genes of interest. While previous cis-regulatory studies in cavefish focused on highly differentially accessible regulatory regions within 10kb of the genes [31], Hi-C data now enables us to expand our understanding to include long-range regulators of gene expression. By integrating looping variations with gene expression data, chromatin accessibility, and epigenetic marks, we can predict potential long-range regulatory elements, as demonstrated in the case of *Arhgef19* with a 40kb long loop.

By comparing morphs of the same species, which evolved traits that are conspicuously different from their ancestral morph within a relative short period of 200,000 years, we can explore how genome architecture may shape evolutionary pathways through physical constraints in the genome. Notably, genome looping anchor points are already associated with genomic instability and recombination hotspots [48]. On the other hand, Quantitative Trait Loci (QTL) analysis provides a list of candidate genomic regions associated with new adaptive traits. Therefore, integrating the Hi-C dataset – which includes information on loop anchors, genomic inversion points, and A/B

boundaries – with QTL markers could offer valuable insights into potential hotspots driving trait evolution in cavefish.

Taken together, this study explores variations in 3D genome architecture from an evolutionary perspective. We highlight the potential role of 3D genome organization in regulating the adaptation of cavefish to the extreme metabolic conditions of cave environments, identifying candidate genes and 3D features associated with metabolic resilience. Additionally, this research lays the foundation for future studies into the 3D genomic basis of extreme adaptation in other organisms.

**Acknowledgement**

The authors would like to thank Kate Hall and Amanda Lawlor from the Stowers Sequencing and Discovery Genomics team for all the help with Library Preparation and Sequencing. We would also like to acknowledge the University of Kansas Medical Center's Genomics Core for their support in generating data on the Illumina NovaSeq 6000 System. The core is supported by the following grants: Kansas Intellectual and Developmental Disabilities Research Center (NIH U54 HD 090216), the Molecular Regulation of Cell Development and Differentiation – COBRE (P30 GM122731-03) and the NIH S10 High-End Instrumentation Grant (NIH S10OD021743). This study is supported by Stowers Institute for Medical Research's institutional funding, and NIH New Innovator Award 1DP2AG071466-01, NIH R24OD030214.

**Author contribution**

**Tathagata Biswas**: Conceptualization; data collection; investigation; analysis; visualization; original draft preparation. **Hua Li**: Analysis; original draft preparation, visualization; data curation. **Nicolas Rohner**: Conceptualization; original draft preparation; funding acquisition.

## Competing interest statement

The authors declare no conflict of interest.

## Data availability statement

Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at http://www.stowers.org/research/publications/libpb-2508

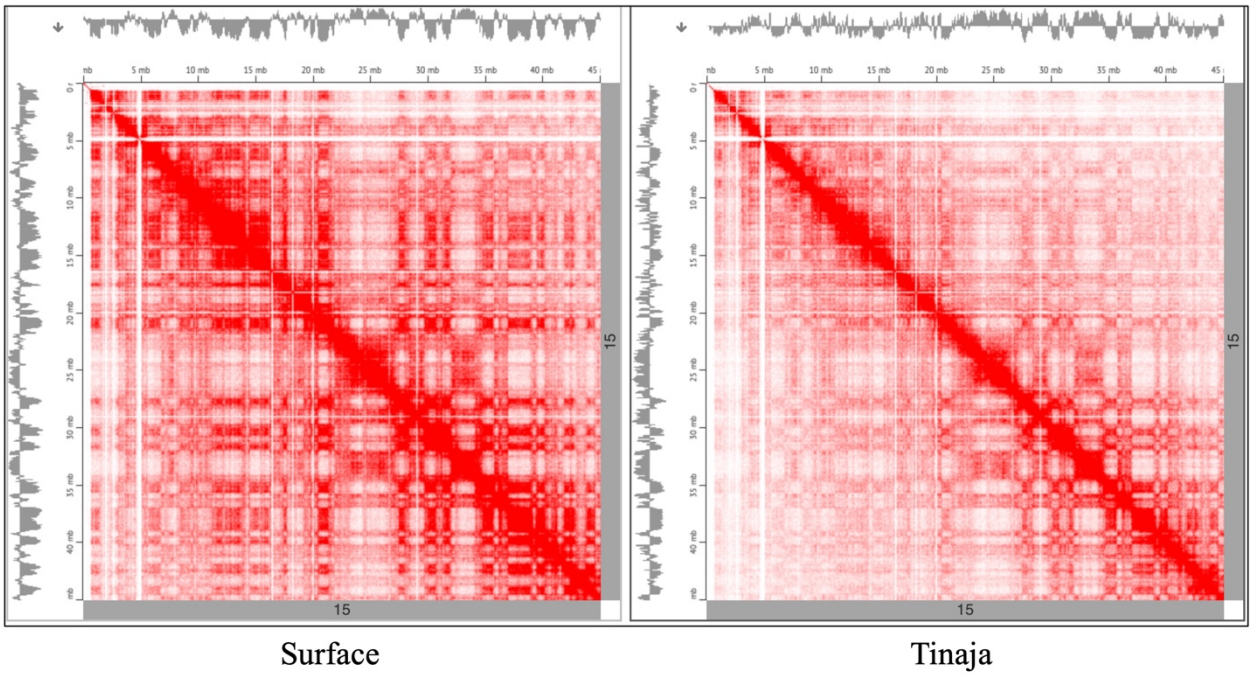
## References


1. Benabdallah, N.S. et al. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell* **76**, 473-484 e477 (2019).
2. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455 (2019).
3. Uyehara, C.M. & Apostolou, E. 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Rep* **42**, 112068 (2023).
4. Golfier, S., Quail, T., Kimura, H. & Brugues, J. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *Elife* **9** (2020).
5. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L.A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A* **115**, E6697-E6706 (2018).
6. Zhang, H. et al. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol Cell* **83**, 2856-2871 e2858 (2023).
7. Batut, P.J. et al. Genome organization controls transcriptional dynamics during development. *Science* **375**, 566-570 (2022).
8. Levo, M. et al. Transcriptional coupling of distant regulatory genes in living embryos. *Nature* **605**, 754-760 (2022).
9. Akdemir, K.C. et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* **52**, 294-305 (2020).
10. Lupianez, D.G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
11. Lupianez, D.G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237 (2016).
12. Despang, A. et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* **51**, 1263-1271 (2019).
13. Williamson, I. et al. Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146** (2019).

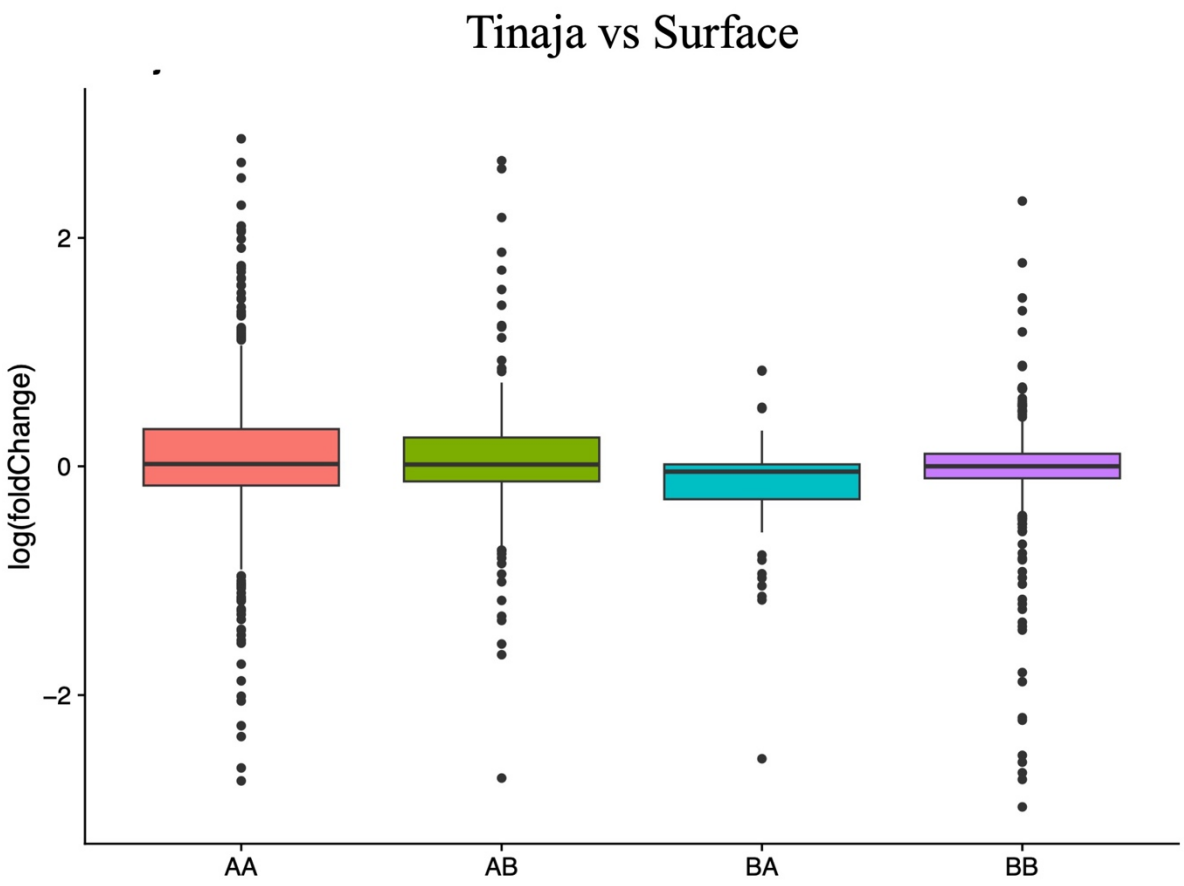
14. Lukyanchikova, V. et al. Anopheles mosquitoes reveal new principles of 3D genome organization in insects. *Nat Commun* **13**, 1960 (2022).
15. Sandoval-Velasco, M. et al. Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample. *Cell* **187**, 3541-3562 e3551 (2024).
16. Yang, Y., Zhang, Y., Ren, B., Dixon, J.R. & Ma, J. Comparing 3D Genome Organization in Multiple Species Using Phylo-HMRF. *Cell Syst* **8**, 494-505 e414 (2019).
17. Fumey, J. et al. Evidence for late Pleistocene origin of Astyanax mexicanus cavefish. *BMC Evol Biol* **18**, 43 (2018).
18. Herman, A. et al. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, Astyanax mexicanus. *Mol Ecol* **27**, 4397-4416 (2018).
19. 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* **112**, 9668-9673 (2015).
20. Medley, J.K. et al. The metabolome of Mexican cavefish shows a convergent signature highlighting sugar, antioxidant, and Ageing-Related metabolites. *Elife* **11** (2022).
21. Pozo-Morales, M. et al. Starvation-resistant cavefish reveal conserved mechanisms of starvation-induced hepatic lipotoxicity. *Life Sci Alliance* **7** (2024).
22. Olsen, L. et al. Metabolic reprogramming underlies cavefish muscular endurance despite loss of muscle mass and contractility. *Proc Natl Acad Sci U S A* **120**, e2204427120 (2023).
23. Riddle, M.R. et al. Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. *Nature* **555**, 647-651 (2018).
24. Xiong, S. et al. Enhanced lipogenesis through Ppargamma helps cavefish adapt to food scarcity. *Curr Biol* **32**, 2272-2280 e2276 (2022).
25. Yamamoto, Y. & Jeffery, W.R. Central role for the lens in cave fish eye degeneration. *Science* **289**, 631-633 (2000).
26. Durand, N.C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
27. Harris, H.L. et al. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat Commun* **14**, 3303 (2023).
28. Shumate, A. & Salzberg, S.L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643 (2021).
29. Ritchie, M.E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
30. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).
31. Krishnan, J. et al. Genome-wide analysis of cis-regulatory changes underlying metabolic adaptation of cavefish. *Nat Genet* **54**, 684-693 (2022).
32. Biswas, T., Hassan, H. & Rohner, N. Potential role of microRNAs in regulating transcriptional profile, and sculpting development and metabolism in cavefish. *bioRxiv*, 2024.2001.2030.578051 (2024).
33. Rowley, M.J. et al. Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals. *Genome Res* **30**, 447-458 (2020).
34. Biswas, T. et al. 3D spheroid culturing of Astyanax mexicanus liver-derived cell lines recapitulates distinct transcriptomic and metabolic states of in vivo tissue environment. *J Exp Zool B Mol Dev Evol* (2024).

35. van der Weele, C.M. & Jeffery, W.R. Cavefish cope with environmental hypoxia by developing more erythrocytes and overexpression of hypoxia-inducible genes. *Elife* **11** (2022).
36. Grubert, F. et al. Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737-743 (2020).
37. Bond, M.L. et al. Chromatin loop dynamics during cellular differentiation are associated with changes to both anchor and internal regulatory features. *Genome Res* **33**, 1258-1268 (2023).
38. Kumar, D., Cinghu, S., Oldfield, A.J., Yang, P. & Jothi, R. Decoding the function of bivalent chromatin in development and cancer. *Genome Res* **31**, 2170-2184 (2021).
39. Zhou, V.W., Goren, A. & Bernstein, B.E. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18 (2011).
40. Horii, T., Morita, S., Kimura, M. & Hatada, I. Epigenetic regulation of adipocyte differentiation by a Rho guanine nucleotide exchange factor, WGEF. *PLoS One* **4**, e5809 (2009).
41. Seki, Y. et al. In Utero Exposure to a High-Fat Diet Programs Hepatic Hypermethylation and Gene Dysregulation and Development of Metabolic Syndrome in Male Mice. *Endocrinology* **158**, 2860-2872 (2017).
42. Wang, W. et al. Cytoplasmic Endonuclease G promotes nonalcoholic fatty liver disease via mTORC2-AKT-ACLY and endoplasmic reticulum stress. *Nat Commun* **14**, 6201 (2023).
43. Warren, W.C. et al. Astyanax mexicanus surface and cavefish chromosome-scale assemblies for trait variation discovery. *G3 (Bethesda)* **14** (2024).
44. Hager, E.R. et al. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**, 399-405 (2022).
45. Harringmeyer, O.S. & Hoekstra, H.E. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat Ecol Evol* **6**, 1965-1979 (2022).
46. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G. & Higgs, D.R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* **26**, 2041-2051 (2007).
47. Krishnan, J. et al. Liver-derived cell lines from cavefish Astyanax mexicanus as an in vitro model for studying metabolic adaptation. *Sci Rep* **12**, 10115 (2022).
48. Krefting, J., Andrade-Navarro, M.A. & Ibn-Salem, J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* **16**, 87 (2018).

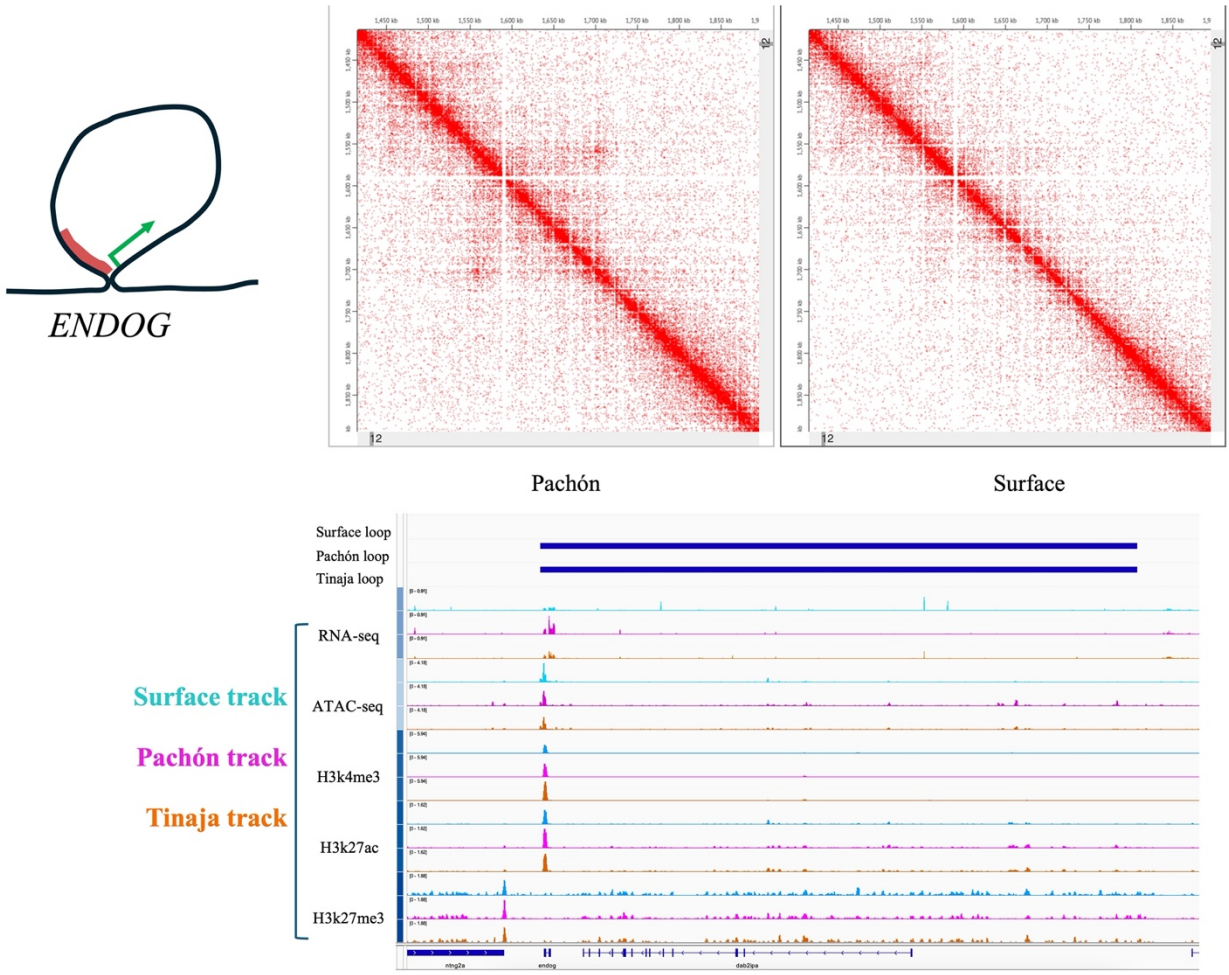
**Supplementary Figures**




Supplementary Fig. 1: Plaid-like compartment and A/B compartment track for chromosome 15 in Surface and Tinaja.




Supplementary Fig. 2: Boxplot depicting fold change in Tinaja, for the genes in the loci which switched compartments. AA and BB denotes unchanged compartments, AB denotes loci which is active in Tinaja and inactive in surface, BA denotes loci active in surface and inactive in Tinaja.




Supplementary Fig. 3: Cave-specific loop, with *Endog* at one of the anchor points. Depicted in Pachón and surface on chromosome 12, integrated with RNA-seq, ATAC-seq and ChIP-seq data.