Divergent 3D genome organization links to metabolic adaptation in cavefish.

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

Affiliation

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

[2] Present address: Institute for integrative cell biology and physiology, University of Münster, Münster, Germany

*Email for correspondence: nro@stowers.org

Abstract

The cave morphs of *Astyanax mexicanus* adapt to caves through distinct metabolic and morphological traits. While changed gene expression is a primary driver of these adaptations, the underlying role of 3D genome organization – a key regulator of gene expression – remains unexplored. We analyzed liver 3D genome architecture in surface and cave morphs and identified cave-specific 3D genome signatures which, when integrated with transcriptomic and epigenetic data, associated with expression changes in genes associated with cave metabolism. In this study, comparing morphs within the same species, we established a foundation for better understanding of how 3D genome may drive phenotypic diversity.

Main

The Mexican tetra, *Astyanax mexicanus*, is a remarkable species that exists in two distinct forms: river-dwelling surface fish and cave-dwelling cavefish, the latter of which have evolved separately across multiple cave populations. These cavefish have adapted to the dark, nutrient-scarce environments of caves, developing striking traits such as the loss of eyes and pigmentation, alongside metabolic adaptations like resilience to starvation, enhanced fat storage (lipogenesis), persistent high blood sugar (hyperglycemia), and excessive eating (hyperphagia) [1-3]. These differences between surface fish and cavefish are largely driven by changes in gene expression, as revealed by comparative molecular studies. While prior research has explored the contributions of cis- and trans-regulatory elements to these metabolic and transcriptomic shifts [4, 5], the role of three-dimensional (3D) genome organization – a critical regulator of gene expression – has remained unexplored.

Within the nucleus, 3D genome organization governs the transcriptomic profile by determining the spatial proximity of regulatory elements, such as enhancers, to their target genes. This organization involves genome looping, where DNA segments spanning a few to hundreds of kilobases fold to bring distant regulatory elements into contact with genes, a process facilitated by cohesin and CTCF-associated protein complexes through loop extrusion [6-8]. The genome is further categorized into active (A) and inactive (B) compartments based on transcriptional activity, gene density, and looping interactions [9].

The stability of 3D genome organization across evolutionary timescales is well-documented, with conserved patterns often linked to consistent gene expression across species. However, species-

specific architectural variations also exist, explaining some inter-species differences in gene expression [10-12]. Yet, little is known about whether different morphs of the same species, evolving under divergent environmental pressures, exhibit distinct 3D genome organizations. If such differences exist, they could connect changes in genome architecture to specific phenotypic adaptations, building on the established role of 3D organization in gene regulation. To explore this, we investigated the 3D genome architecture of *A. mexicanus* surface fish and two cavefish morphs – Pachón and Tinaja – focusing on the liver due to its central metabolic role and the availability of extensive transcriptomic and epigenetic datasets (Fig. 1a). Using Hi-C, a technique for mapping genome-wide chromatin interactions, we generated the first 3D genome contact maps for this species.

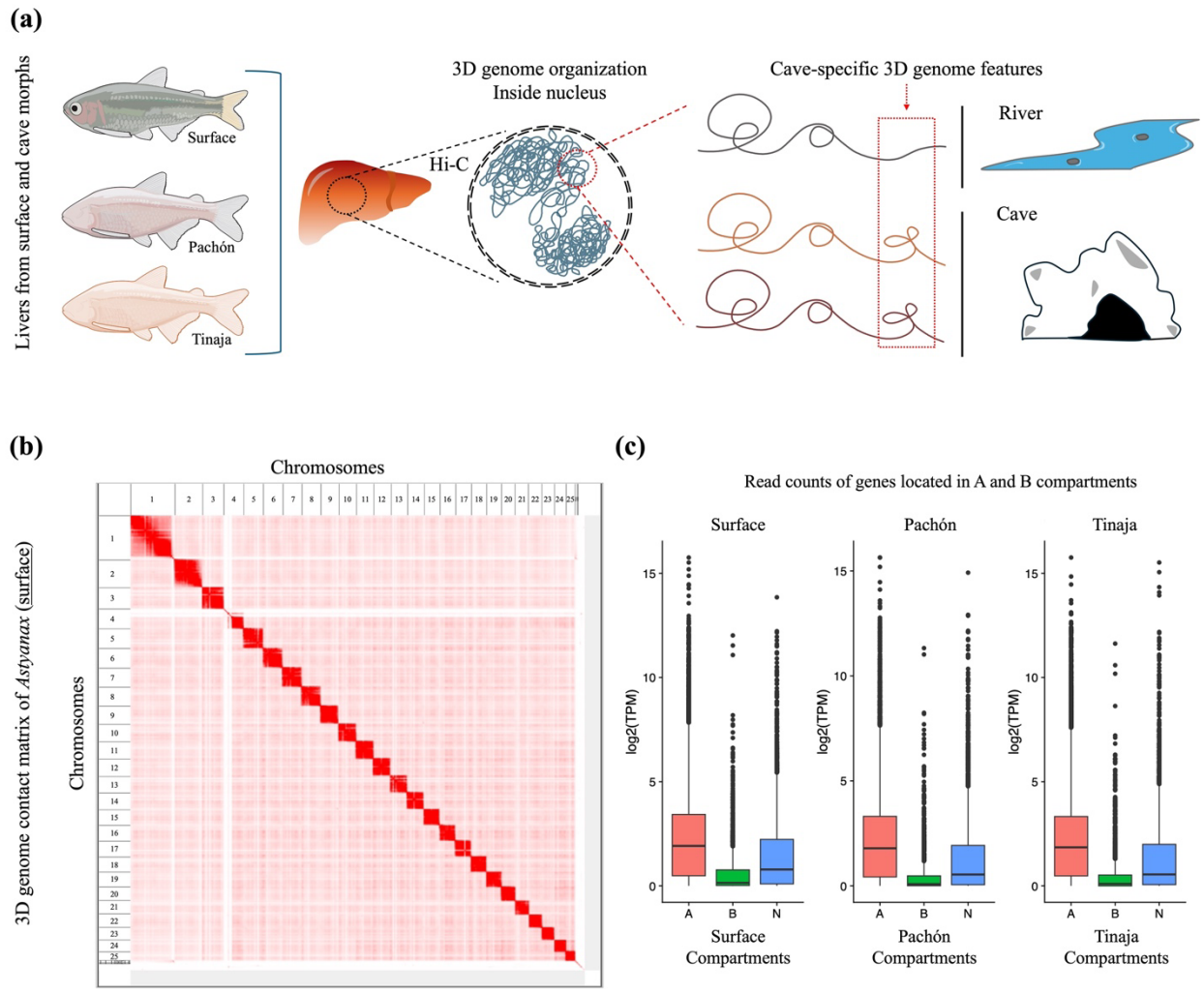


Figure 1: (a) Hi-C experimental plan on livers of *A. mexicanus*. (b) Whole genome 3D contact matrix of *A. mexicanus* (surface) liver derived from Hi-C data. (c) Gene residing in the active 'A' compartment show higher transcriptional activity across all the three morphs of *A. mexicanus* as compared to genes residing in the 'B' compartment. N being the neutral compartment.

We generated Hi-C-based contact matrices for the livers of surface fish, Pachón, and Tinaja, isolating liver tissues from adults and following the Arima-HiC protocol. After sequencing and merging biological replicates, we obtained 1,348 million valid Hi-C reads for surface fish, 1,093 million for Pachón, and 603 million for Tinaja, averaging 304 million contacts per morph. These data enabled high-resolution contact matrices (up to 5 kb), revealing clear delineation of the 25 chromosomes in each morph, with strong intra-chromosomal interactions (Fig. 1b).

Next, we analyzed the segregation of the genome into A and B compartments, a hallmark of 3D chromatin organization observed across species. Genomic loci with higher gene density and transcriptional activity tend to cluster in the 'A' compartment, while less active regions segregate into the 'B' compartment. Using Principal Component Analysis (PCA) Of Sparse, SUper Massive Matrices (POSSUMM) [13], we identified these compartments in all three morphs (Fig. S1), confirming that the A compartment housed genes with higher transcriptional activity than the B compartment, as validated by mapping liver RNA-seq data (Fig. 1c). However, adaptation to the cave environment, known to induce large-scale transcriptomic and regulatory shifts [4, 5], prompted us to investigate whether compartmental changes contributed to these differences. Comparing the morphs, we found that while compartmentalization was largely conserved, 665 out of 53,988 genomic bins (25 kb each) exhibited A/B switching in cavefish relative to surface fish ($p \leq 0.001$). Genes in regions switching to the A compartment in cavefish showed higher transcriptional fold changes than those switching to the B compartment (Fig. S1), suggesting that large-scale chromatin reorganization supports transcriptomic shifts during cave adaptation.

To test whether these compartmental switches were linked to cave adaptation, we sought switches common to both Pachón and Tinaja but absent in surface fish – termed cave-specific genome compartmental switches. We identified regions switching from B in surface fish to A in both cave morphs (BAA regions), containing 287 genes, and from A to B (ABB regions), containing 36 genes (Supp. Table 1). Gene Ontology (GO) analysis of BAA genes revealed presence of genes that contribute to pyrimidine-related pathways, which was consistent with prior findings that intermediates like orotic acid respond to starvation in *A. mexicanus* liver tissue [14].

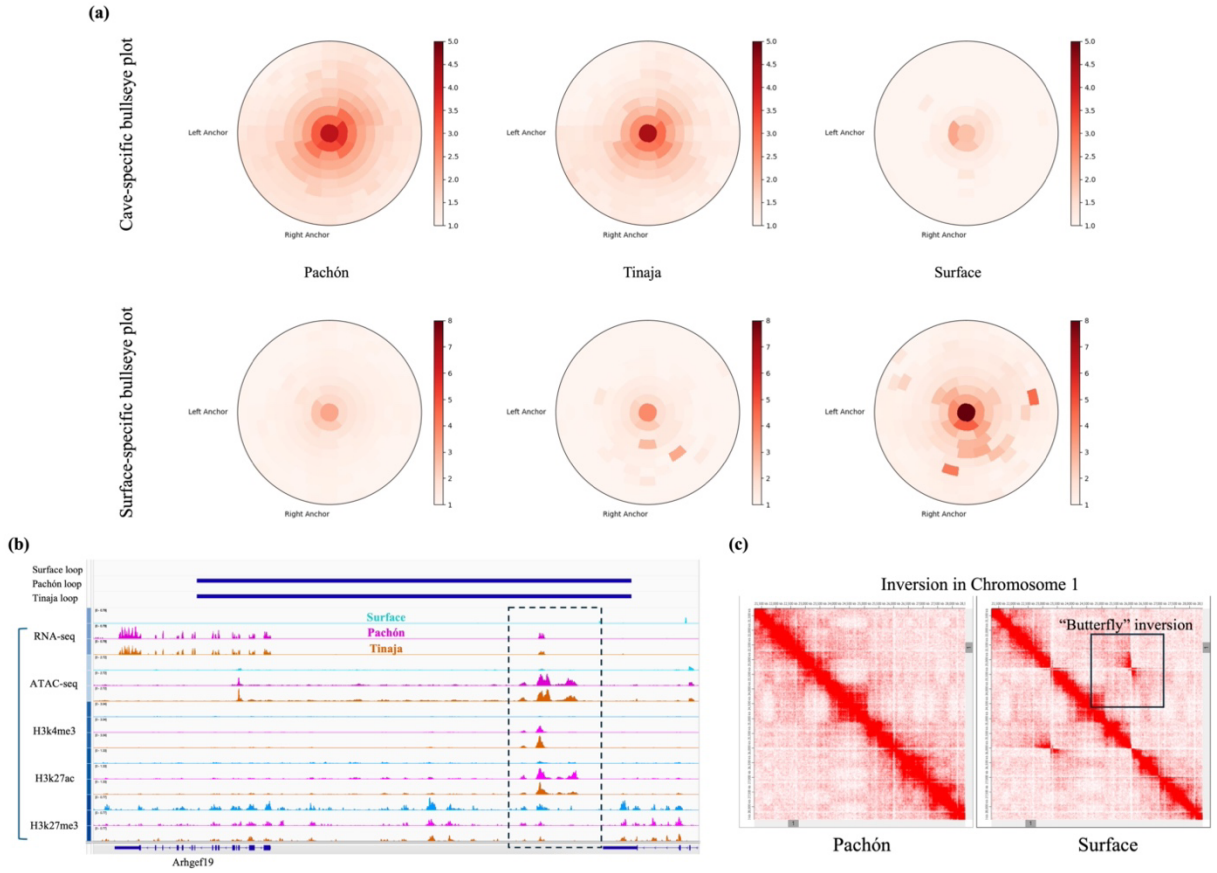


Figure 2: (a) Bullseye transformed visualization from SIPMeta plots for cave- (top) and surface-specific (bottom) loops. (b) Aligning the RNA-seq, ATAC-seq and signature histone ChIP-seq data with the loop calls, to observe activity around the loop anchor points. The box marks the putative regulatory region near the other anchor point of the 40kb *Arhgef19* loop. (c) Genomic inversion in chromosome 1 ("butterfly" pattern) in highlighted in the dark box. To be noted that the inversion appears on surface contact matrix, since the reference is Pachón genome (AMEX_1.1).

At a finer scale, we examined chromatin looping, which facilitates enhancer-promoter interactions critical for gene regulation. Using the SIP caller [15], we identified 2,525 loops in surface fish, 3,459 in Pachón, and 1,690 in Tinaja, with ~1,500 loops shared across all morphs. However, 232 loops were unique to surface fish, and ~200 were common to both cave morphs but absent in surface fish, being cave-specific. Bullseye metaplots confirmed strong morph-specific signal enrichment at these loops, suggesting regulatory interactions unique to each environment (Fig. 2a). However, despite these differences, putative CTCF motif orientations at the loop anchors remained

conserved, predominantly convergent, indicating stable loop stabilization mechanisms across morphs (Fig. S2).

Thereafter, integrating the Hi-C data with RNA-seq, ATAC-seq, and histone ChIP-seq datasets, we assessed the transcriptional impact of these loops. Of 620 genes in cave-specific loops, 117 were differentially expressed ($p \leq 0.01$), with 75 upregulated in cavefish (Supp. Table 2), while 146 of 667 genes in surface-specific loops were differentially expressed, with 102 upregulated in surface fish (Supp. Table 2). Notably, *Arhgef19*, with the highest fold change among these genes, was near a cave-specific loop anchor and linked to an active enhancer near the other anchor point in the loop (marked by H3K27ac and ATAC-seq peaks), suggesting a cave-specific regulatory mechanism (Chr 13:36,235,001-36,275,000) (Fig. 2b). Interestingly, *Arhgef19* has previously been associated with adipocyte differentiation and metabolic response to high-fat diets [16, 17]. Furthermore, in a starvation study, *Arhgef19* was differentially expressed in the livers of zebrafish subjected to starvation compared to those that were starved but subsequently rescued from liver steatosis and atrophy [18].

Finally, Hi-C analysis using the EagleC pipeline [19], also identified genomic inversions, detectable as 'butterfly' patterns. Inversions are known to drive fitness differences across ecotypes adapting to distinct environments [20]. Notably, a prominent ~2.7 Mb inversion on chromosome 1 (Chr 1:23,310,000-26,040,000), shared by both cave morphs (Fig. 2c), harbors genes associated with cell cycle regulation (Supp. Table 3). Given the role of cell cycle control in resilience to physiological stress, this inversion may contribute to metabolic adaptations in cavefish.

In conclusion, our study reveals significant 3D genome differences between *A. mexicanus* surface and cave morphs, with cave-specific features putatively linked to metabolic adaptation. By identifying candidate genes like *Arhgef19* through loops analysis, and structural features like inversions, we provide novel insights into the genomic basis of cavefish adaptation, setting the stage for future research into 3D genome mechanisms in extreme environments. It should also be noted, that while many genes in loci of altered 3D organization showed no immediate transcriptional change, these architectural changes may prime regulatory elements for rapid responses to stress, enhancing adaptability. Future *in vivo* and *in vitro* [21, 22] metabolic assays targeting candidate genes within these altered loci will help elucidate their role in adaptability. Unlike most 3D genome studies, which compare different species, our work examines morphs within a single species adapted to vastly different environments, laying a foundation for better understanding of how 3D genome changes contribute to phenotypic diversity. This work not only advances our understanding of *A. mexicanus* evolution but also highlights the broader significance of genome architecture in phenotypic diversity within a single species.

Methods

Sample collection for Hi-C

Liver tissues were collected from three adult fish from each morph: Pachón (499 days, 1M, 2F), Tinaja (511 days, 1M, 2F), and surface fish (553 days, 3F). 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 oscillations/minute for 30secs. Thereafter, Hi-C data was generated using the Arima-HiC kit (PN: A510008), which streamlined

the protocol as it included reagents optimized for efficient chromatin digestion, proximity ligation, and DNA purification. After necessary QC, sample were ready for library preparation.

Libraries and Sequencing

The Pachón and surface fish data were generated in the first round. Hi-C libraries were prepared 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). Libraries were pooled, re-quantified and sequenced as 150bp paired reads on an S1 flow cell (v1.5) on the Illumina NovaSeq 6000 instrument (Surface and Pachón) and P3 flow cell on the Illumina Nextseq 2000 instrument (Tinaja), utilizing RTA and instrument software versions current at the time of processing.

HiC-seq data process

The Hi-C experiment reads were processed using Juicer from Aiden lab with the default parameters [23], 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 [13]. The choice of the first or the second PCA (Principal Component Analysis) depended on how it corresponded to the plaid-like pattern observed in the Hi-C data. Once the PCA was decided, the signs of these values were determined to make sure that the positive values always correspond to the active state, which is represented by higher gene density as compared to the inactive state. The significant A/B bin switches were identified using lm() function in base R package. We intersected the gene model with A/B bins using bedtools intersect.

Loop analysis using SIP method and CTCF motifs

Individual sample data (each .hic file) within each morph were merged to form one dataset (one merged .hic file) for each of the three *Astyanax* morphs. We used the SIP method to detect the genomic loops form the data. 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. Loops that are wider than 1MB were excluded. Additionally, MotifFinder was used to find putative CTCF motifs in the anchor points of the identified 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 [24] 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 [25].

Data availability

The Hi-C data sets has been uploaded to GEO database with accession number GSE290882 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE290882). Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at http://www.stowers.org/research/publications/libpb-2508.

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.

Contributions

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 interests

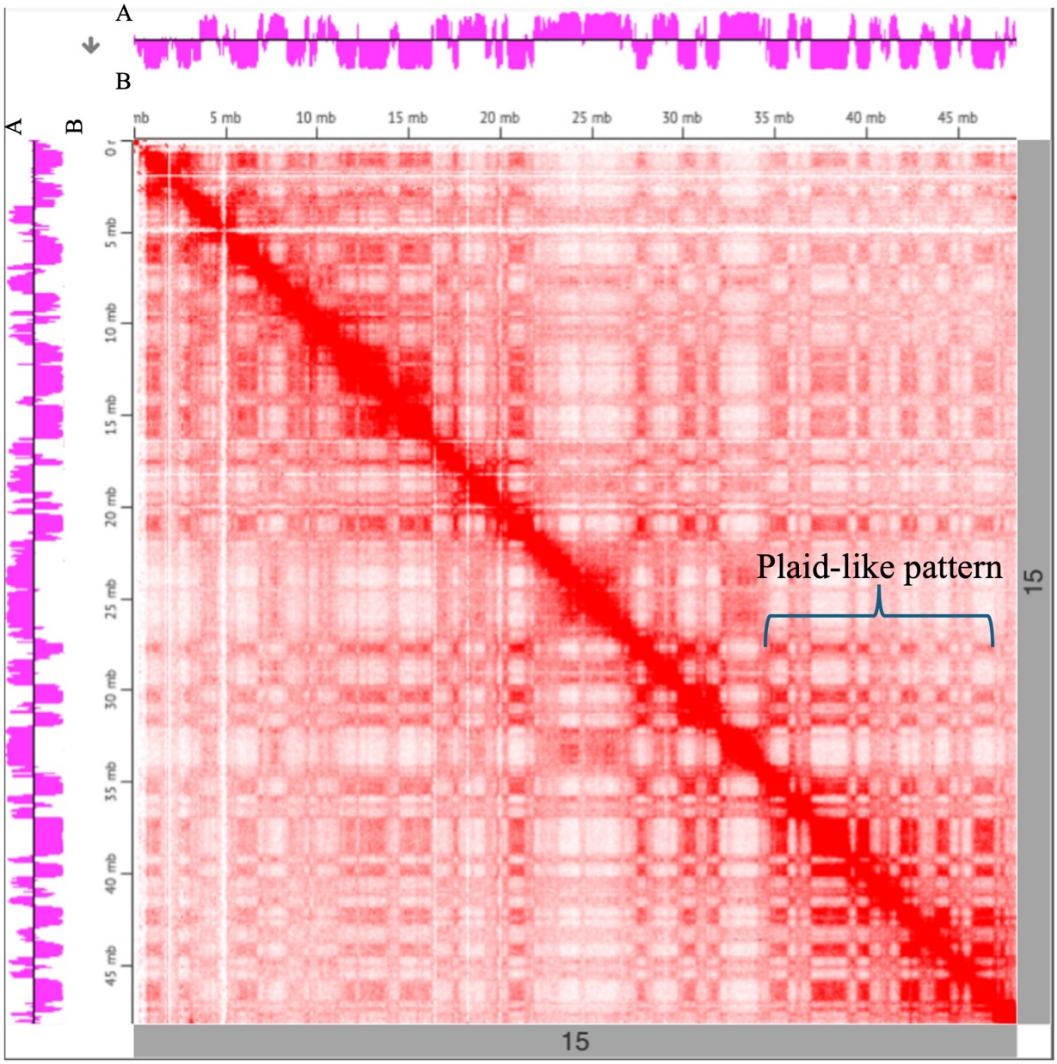
The authors declare no conflict of interest.

References

1. 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).
2. Riddle, M.R. et al. Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. *Nature* **555**, 647-651 (2018).
3. Xiong, S. et al. Enhanced lipogenesis through Ppargamma helps cavefish adapt to food scarcity. *Curr Biol* **32**, 2272-2280 e2276 (2022).
4. Biswas, T., Hassan, H. & Rohner, N. Differentially expressed miRNAs offer new perspective into cave adaptation of Astyanax mexicanus. *Ann N Y Acad Sci* (2025).
5. Krishnan, J. et al. Genome-wide analysis of cis-regulatory changes underlying metabolic adaptation of cavefish. *Nat Genet* **54**, 684-693 (2022).
6. Golfier, S., Quail, T., Kimura, H. & Brugues, J. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *Elife* **9** (2020).
7. 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).
8. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455 (2019).
9. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

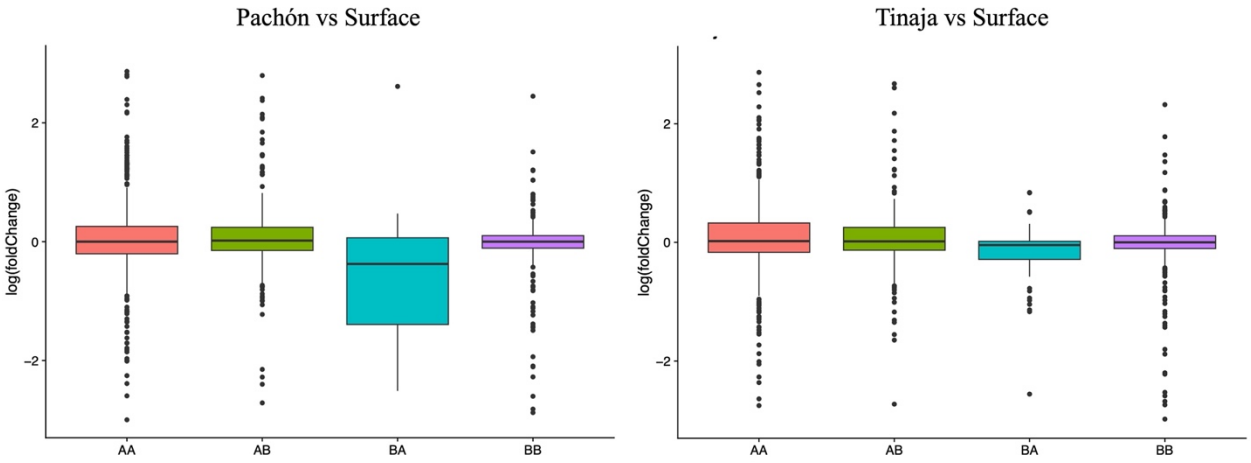
10. Krefting, J., Andrade-Navarro, M.A. & Ibn-Salem, J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* **16**, 87 (2018).
11. Lukyanchikova, V. et al. Anopheles mosquitoes reveal new principles of 3D genome organization in insects. *Nat Commun* **13**, 1960 (2022).
12. 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).
13. Harris, H.L. et al. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat Commun* **14**, 3303 (2023).
14. Medley, J.K. et al. The metabolome of Mexican cavefish shows a convergent signature highlighting sugar, antioxidant, and Ageing-Related metabolites. *Elife* **11** (2022).
15. 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).
16. 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).
17. 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).
18. Pozo-Morales, M. et al. Starvation-resistant cavefish reveal conserved mechanisms of starvation-induced hepatic lipotoxicity. *Life Sci Alliance* **7** (2024).
19. Wang, X., Luan, Y. & Yue, F. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. *Sci Adv* **8**, eabn9215 (2022).
20. Hager, E.R. et al. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**, 399-405 (2022).
21. 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).
22. 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).
23. Durand, N.C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
24. Shumate, A. & Salzberg, S.L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643 (2021).
25. Ritchie, M.E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

Supplementary Figures



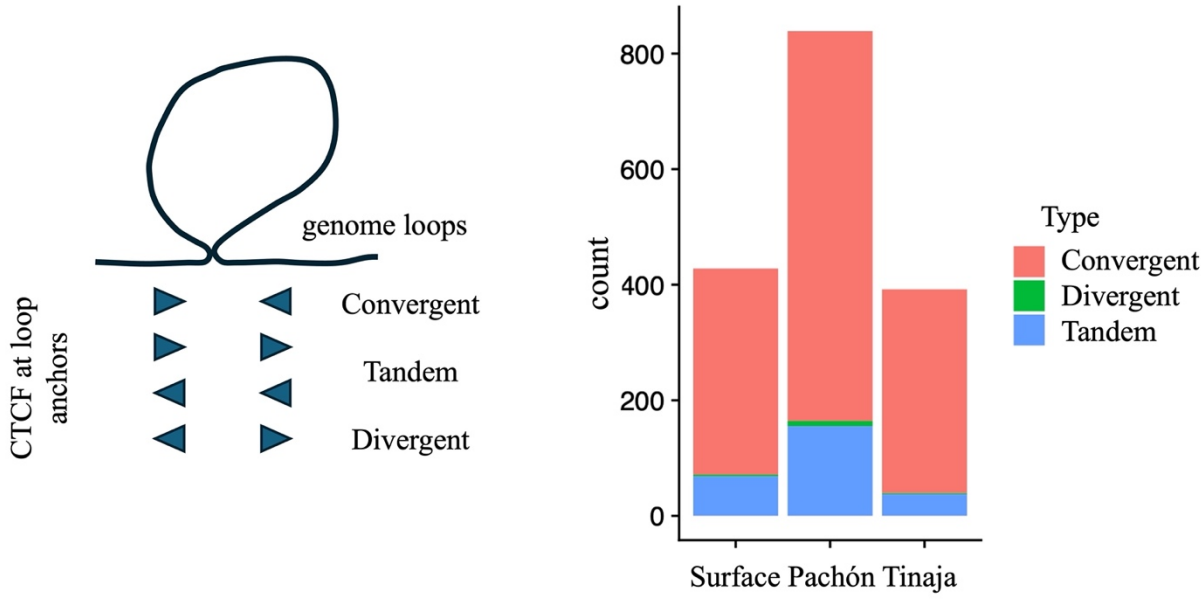
A/B genome compartments (Chromosome 15)

Supplementary Figure S1: Plaid-like patterns in the chromosomal interaction matrix (chromosome 15) are observed in Pachón, indicating A/B compartmentalization. The magenta track adjacent to the axes represents A/B compartments, aligning with the plaid-like pattern. Positive values correspond to A compartments (active, gene-rich regions), while negative values indicate B compartments (inactive, gene-poor regions).



Fold change in gene expression, for genes in the loci which switched compartments.

Supplementary Figure S2: Boxplot depicting fold change in Pachón and Tinaja, for the genes in the loci which switched compartments. AA and BB denotes unchanged compartments, AB denotes loci which is active in caves and inactive in surface, BA denotes loci active in surface and inactive in caves.



Supplementary Figure S3: Distribution of CTCF anchor points for loops in each of the three morphs. The CTCF sites were either convergent (Surface: 330; Pachón: 677; Tinaja: 456 loops), in tandem (Surface: 55; Pachón: 167; Tinaja: 47 loops) or divergent (Surface: 2; Pachón: 12; Tinaja: 3 loops), as depicted in the cartoon.